Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 5, 1065–1081

Copyright 2004 by the American Psychological Association
0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.5.1065

Modeling the Effects of Prior Knowledge on Learning Incongruent Features of Category Members

Evan Heit
University of Warwick

Janet Briggs
University of Kent

Lewis Bott
New York University

The authors conducted 3 experiments addressing the issue of how observations and multiple sources of prior knowledge are put together in category learning. In Experiments 1 and 2, learning was faster for critical features, which were predictable on the basis of prior knowledge, than for filler features, and this advantage increased as more observations were made. In addition, learning was fastest for incongruent features that could only be predicted using knowledge from other domains. In Experiment 3, presenting contradictory features that violated prior knowledge led to rote learning rather than use of prior knowledge. The results were simulated with the Baywatch model, which addresses how observations of category members lead to recruitment and selection of sources of prior knowledge.

Perhaps the most crucial point about category learning is that it does not take place in isolation. When people learn about a category, there is a complex interplay between their observations and their expectations. Prior knowledge shapes category learning. For example, suppose that a new type of breakfast cereal is introduced in supermarkets. Because people already have extensive prior knowledge about breakfast cereals and other foods, their direct observations of this cereal will comprise a relatively small proportion of their knowledge base compared with their prior knowledge.

Before introducing our own experiments, which focused on how prior knowledge affects the learning of unexpected information in categories, we provide some background in terms of theoretical and experimental work in this area. There have been many theoretical arguments for (e.g., Murphy & Medin, 1985) and experimental demonstrations of (e.g., Hayes, Foster, & Gadd, 2003; Heit, 1994, 1998b, 2001b; Kaplan & Murphy, 2000; Murphy & Kaplan, 2000; Palmeri & Blalock; 2000; Wisniewski & Medin, 1994) the pervasive effects of prior knowledge on category learning (see Heit, 2001a, and Murphy, 2002, for reviews). One important generalization is that prior knowledge facilitates the learning of information that is congruent with expectations. For example, Kaplan and Murphy (2000) compared learning about contrasting categories, for which each category member had a feature consistent with some theme (e.g., underwater buildings vs. space buildings), with learning about categories that did not consistently follow these contrasting themes. It was found that category learning was faster, and classification of features was more accurate, when the categories were consistent with expectations.

What is important about this phenomenon of facilitation due to prior knowledge is that it is potentially informative for the development of theoretical accounts of category learning. That is, determining when facilitation does and does not take place and what kinds of knowledge have this effect can be used to develop more complete models of category learning.

Heit and Bott (2000) showed that the phenomenon of facilitation due to prior knowledge itself varies in magnitude over the course of category learning. With so much prior knowledge available, sometimes the benefits of prior knowledge will not be manifest until enough data have been collected to select from the many possible sources of prior knowledge. Hence, knowledge selection is a key part of category learning. In Heit and Bott's study, participants learned about two kinds of buildings, Doe and Lee. They were shown a series of descriptions, each corresponding to a different building. The stimuli were organized in five blocks, with descriptions of four Doe buildings and four Lee buildings presented in each block. Each description included the category label, Doe or Lee, and a list of featural information. There were two critical features presented in each description and two filler features. The critical features for each category were related to a familiar type of building, for example, churches for Doe and office blocks for Lee. Hence, the critical features were congruent with prior knowledge abut category subtypes. The filler features, arbitrarily assigned to each category, were general characteristics that could be true of just about any building. For example, *lit with candles* was a critical feature, and *near a river* was a filler feature.

Evan Heit, Department of Psychology, University of Warwick, Coventry, United Kingdom; Janet Briggs, Department of Psychology, University of Kent, Canterbury, United Kingdom; Lewis Bott, Department of Psychology, New York University.

These experiments were conducted as part of a research project funded by a grant from Unilever. We are grateful to Gordon Bower, Paula Durlach, Nieke Gerritsen, Hilde Hendrickx, Stephan Lewandowsky, Gregory Murphy, Robert Nosofsky, and Debbie Slater for comments on this research.

Correspondence concerning this article should be addressed to Evan Heit, Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom. E-mail: e.heit@warwick.ac.uk

The main result was that critical features were learned better than filler features, but this advantage of critical features depended on the training block. In the first training block, when participants had little data to use to select from sources of prior knowledge, there was no advantage for critical features. However, with more training blocks, there was an increased advantage for critical features. In other words, at the start of the experiment, people had far too much prior knowledge about buildings for it to be of any use. However, with more observations, it became clear that prior knowledge about churches and office buildings would be particularly useful.

Heit and Bott (2000) presented Baywatch, a simple connectionist network model, as an account of these results. The model's name refers to its combination of a general Bayesian approach to selecting among multiple sources of prior knowledge (e.g., Heit, 1998a; Tenenbaum & Griffiths, 2001) with an empirical learning component. This approach also has some parallels to mixture-of-experts networks (e.g., Jacobs, 1997; see also Bott & Heit, 2004; Erickson & Kruschke, 1998; Lewandowsky, Kalish, & Griffiths, 2000; Yang & Lewandowsky, 2003), using modules with different pools of pretrained knowledge. The Baywatch model, illustrated in Figure 1, can be described as having one module or set of weights for strictly empirical learning. These weights do not get any pretraining. The model also has a set of experts that are pretrained to recognize different known categories. For example, a network for learning about buildings might have experts that can recognize different kinds of buildings such as churches, office blocks, restaurants, and schools. (Only two of these expert modules are illustrated in Figure 1.) The Baywatch model is a feedforward network in which the input units represent the individual features, and the output units represent the Doe and Lee category nodes. The two hidden units correspond to two expert modules or prior knowledge (PK) nodes. The four input units on the left of Figure 1 represent filler features, and the four inputs on the right represent the critical features. The only difference between the two types of features is that the filler features are only connected to the output nodes, whereas the critical features are connected both directly to the output nodes and indirectly to the output nodes via the PK nodes. The connections between the critical features and the PK nodes have fixed weights; thus, critical features of the stimuli that correspond to church features would activate the church PK node, and likewise critical features of the stimuli that correspond to offices would activate the office PK node. The PK nodes have threshold functions; thus, if any church feature, say, *lit with candles,* is presented, then the church PK node will be activated. The activation from the PK node would then be propagated to the output units. In contrast to the connection weights between the critical features and the PK nodes, the other weights in the network are learnable through gradient descent.

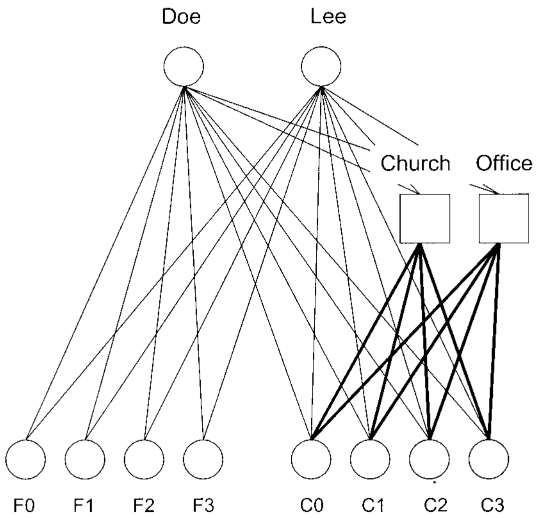

Figure 1. The Baywatch model as applied in Heit and Bott (2000).

The Baywatch model accounted for the key results (see Heit & Bott, 2000, for further details). The model predicted that both critical and filler features will be learned as more observations are made. There would be learning of the direct empirical connections between the input feature nodes and the output nodes for the category labels. Initially, there would be no advantage of critical features over filler features, but with more learning, prior knowledge would be used increasingly to benefit the critical features as the connections between the PK nodes and the output nodes are strengthened. The advantage for critical features over filler features would derive from having two sets of connections for making predictions, on the empirical side and on the knowledge-mediated side.

Rationale of Experiments

When discussing the role of knowledge in guiding category learning, it is important to keep in mind that the same item can belong to multiple categories. Multiple sources of knowledge can be relevant. Borrowing an example from Ross and Murphy (1999), a bagel can be categorized taxonomically (as a type of bread) in terms of ingredients (as a starchy food), as well as by function (as a breakfast food). Indeed, relying on modern technology, foods are often sold as belonging to multiple, seemingly conflicting categories, such as margarines that lower cholesterol or desserts that boost the immune system. Likewise, cross-classification is prevalent for social categories (e.g., Macrae, Bodenhausen, & Milne, 1995), for example, the same person can be a woman, a European, a kick boxer, and a school teacher. What is expected for one category may be unexpected for another category. If prior knowledge shapes category learning, then category learning must be affected by multiple sources of prior knowledge, and there are potential conflicts of expectations.

Whereas Heit and Bott (2000) compared learning about critical features that are expected on the basis of prior knowledge with filler features that are neutral with respect to prior knowledge, the present experiments also examined features that would be unexpected on the basis of prior knowledge. The aim was to further develop the Baywatch model of effects of prior knowledge on category learning to address the use of multiple sources of knowledge. These experiments used a domain, breakfast cereals, for which people would have extensive prior knowledge. In these experiments, participants learned about two new kinds of breakfast cereals labeled Daily and Key. Both kinds of breakfast cereal were described with filler features, such as *box is white* and *made with wheat* that could be true of any cereal. In addition, one kind of cereal had critical features that were linked to prior knowledge about adult or healthy cereals, such as *low in sugar* and *flavored with fruit.* The other kind of cereal had critical features that were linked to prior knowledge about children's or unhealthy cereals, such as *high in sugar* and *artificially flavored.* It was predicted that for these categories, the pattern of learning critical and filler features would be similar to that in the Heit and Bott study.

The major change in the present experiments was the introduction of incongruent features. In Experiments 1 and 2, these incongruent features came from outside the domain of breakfast cereals and hence would be unexpected on the basis of prior knowledge of cereals. For the adult cereal, there were additional incongruent features linked to falling asleep, such as *helps with insomnia.* For the children's cereal, there were additional incongruent features linked to cooked breakfasts, such as *flavored with bacon.* (In Experiment 3, incongruent features of a different type were examined.)

The Baywatch model had not been previously applied to experiments with incongruent information. The most straightforward way to apply the model to Experiments 1 and 2 is shown in Figure 2. Here, the incongruent features at the far left would be learned like filler features. That is, it would be possible to learn rote connections between incongruent features and the category labels, but there would be no mediation due to prior knowledge about breakfast cereals. Hence, incongruent features would be learned more slowly than critical features. Although the incongruent features would not be expected, they would be equally unexpected for adults' and children's cereals. Therefore, the model, as configured, would predict that incongruent features should be at the same level as filler features, which also were equally associated with adults' and children's cereals (see the Appendix for implementation details of this simulation and an illustration of the predictions).

At this point, we only state the prediction for incongruent features on the basis of this baseline application of the Baywatch model. It is clear that other predictions would be possible; however, we defer the discussion of alternative theoretical explanations and predictions until after the results of Experiment 1 are presented.

Experiment 1

This experiment was based on the study of Heit and Bott (2000), with two main changes. First, participants learned about categories of breakfast cereals, coming from the knowledge-rich domain of foods (e.g., Ross and Murphy, 1999). More significantly, in addition to critical features, which were congruent with prior knowledge about types of cereals and filler features, which were neutral with respect to prior knowledge, the categories were presented with incongruent features relating to sleeping pills or cooked breakfasts. Hence, the main purpose of Experiment 1 was to examine the joint effects of observations and prior knowledge on learning about incongruent features.

Method

Subjects. There were 46 paid participants, mainly students, recruited on the University of Warwick campus.

Materials. The participants learned about two categories of breakfast cereals, referred to as Daily and Key cereal brands. The participants were told to imagine that they were reading information from market research interviews that recorded details recalled by respondents after tasting the cereals. The stimuli were presented in six blocks, each block containing descriptions for three samples of Daily cereal and three for Key cereal. Each description consisted of a list of eight features presented in a random order. There were two critical features presented in each description and two filler features. The critical features for each category were related to a familiar cereal type (e.g., children's cereal for Daily and adult cereal for Key or vice versa). The filler features, arbitrarily assigned to each category with a different randomization for each participant, were general characteristics that could be true of any cereal. Each description contained a novel incongruent feature relating to inducing sleep for the adult cereal or cooked breakfasts for the children's cereal. Finally, each description contained three pieces of individuating information (name of shopper, name of interviewer, and use-by date). This information changed with each presentation of the description, serving to make learning somewhat more challenging, as well as emphasizing that each presentation was unique (see also Barsalou, Huttenlocher, & Lamberts, 1998). Results for the individuating features are not reported here.

Each cereal type was assigned eight critical features, distinctive for either children or adults, and eight filler features that would be generic for

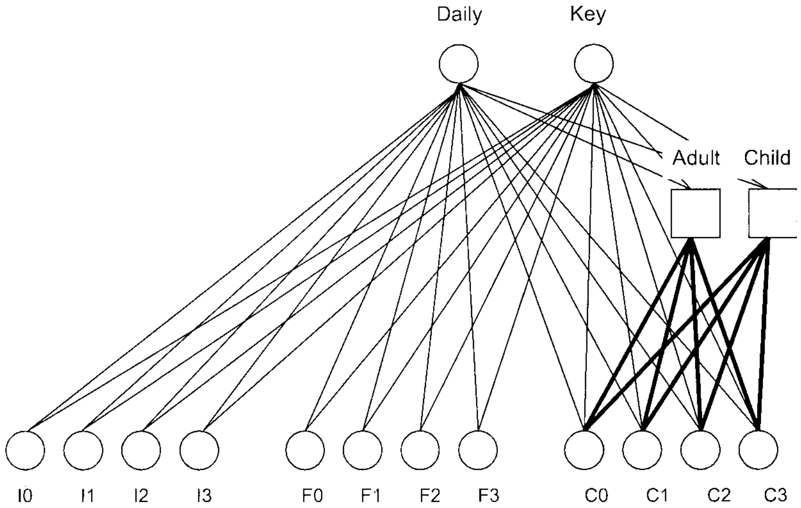

Figure 2. The Baywatch model with incongruent features shown.

most cereals and would not be distinctive for either child or adult cereals. These features are listed in Tables 1 and 2. Note that the features are organized in pairs; thus, one feature, for example, *high in sugar* would apply to one cereal, and its alternative, for example, *low in sugar,* would apply to the other cereal. In addition, there were incongruent features that did not fit usual prior knowledge about breakfast cereals.

The selection of features was informed by pretests, following a similar procedure to Heit and Bott (2000). Initially, 18 pairs of features were prepared: 9 intended to be critical pairs and 9 intended to be filler pairs. The critical pairs were selected to have one feature characteristic of children's breakfast cereals and one feature characteristic of adult breakfast cereals. The filler pairs had two features that could each be equally characteristic of adult or children's cereals. In the first pretest, 15 other participants were instructed to freely sort features into two groups. The sorting task was used to discard critical features if participants did not show a strong preference for putting them in one category, and likewise filler features were discarded if participants did show a strong preference for one category or the other. On this basis, we discarded one critical pair and one filler pair. The remaining critical pairs were all sorted consistently into the same categories by at least 14 of the 15 participants. Of the filler pairs, all but 3 pairs were sorted into the same category by no more than 10 of the 15 participants, and modifications were made to 2 of these 3 pairs before further pretesting.

A second pretest was used on the remaining critical and filler features, as well as features that were intended to be incongruent with prior knowledge of adults' and children's cereals. Eighteen other participants were asked whether each feature is more typical of children's or adults' cereals. For these ratings, ranging from 1 (*more typical of children's cereals*) to 7 (*more typical of adults' cereals*), the ratings for those critical features intended to be typical of adult cereals (range: 5.3–6.2), whereas those for the critical features intended to be typical of children's cereals (range: 1.1–2.7). Mean ratings for the filler features on the adult–child scale ranged from 2.7 to 4.7, and for the incongruent features it ranged from 3.9 to 4.6. Hence, in comparison with the critical features, the filler and incongruent features did not seem to have strong prior associations favoring either child or adult cereals. In a final pretest, we confirmed our intuitions that the incongruent features were considered highly inconsistent with breakfast cereals on the whole.

From these pretests, eight binary pairs of critical features, eight binary pairs of filler features, and two incongruent characteristics, each with three variants, were obtained. One set of incongruent features was *helps you get to sleep, helps with insomnia,* and *promotes relaxation.* The other set of incongruent features was *flavored with bacon, contains scrambled eggs,* and *full English breakfast flavor.*

Procedure. There was a sequence of six study-test blocks. In each study block, the descriptions were presented individually for 6 s each. A sample description would be as follows: {high in sugar, shopper: G. Dean, toy in box, flavored with bacon, green printing on box, interviewer: J. Giddings, made with wheat, use by 5th February 1998}. Participants were instructed to try to learn all of the information presented. For each participant, six critical feature pairs were chosen randomly from the list in Table 1, and six filler feature pairs were chosen from the list in Table 2. These features were presented to participants during study blocks. The remaining critical and filler features were never presented during study blocks.

Following each study block was a brief interleaved task, not reported here, in which participants answered a series of 10 evaluation questions regarding the cereals, such as whether they would like to buy each cereal. Participants were then tested on all of the features presented in the descriptions and on the two critical features and two filler features per category not presented during study. The test consisted of 56 items: 16 critical features, 16 filler features, 6 incongruent features, and 18 individuating features. Of the critical features and the filler features, 12 had been presented during the study blocks and 4 had not. The features were presented individually in a random order, and participants were asked to categorize them as Daily or Key cereal. Overall accuracy feedback was given at the end of each test block to encourage good performance. Stimulus presentation and data collection were carried out using a computer.

Table 1
Critical Features

Child cereal	Adult cereal
High in sugar	Low in sugar
Comes with voucher for amusement park	Comes with voucher for coffee
Animal-shaped	Flake-shaped
Advertised in pop music magazines	Advertised in Sunday newspapers
Cartoon character on box	Photograph on box
Toy in box	Health tips on box
Multicolored	Brown-colored
Artificially flavored	Flavored with fruit

Table 2
Filler Features

Set 1	Set 2
Ingredients listed on side of box	Ingredients listed on back of box
Box is white	Box is cream-colored
Made with wheat	Made with corn
Opaque bag inside box	Clear bag inside box
Sold in 475 g box	Sold in 450 g box
Green printing on box	Blue printing on box
Made in London	Made in Bristol
Costs £1.85	Costs £1.89

Results

The key results are shown in Figure 3 separately for the features that were presented and for the features that were tested but never presented during study. First focusing on the left side, it is clear that for all types of features, accuracy improved with more presentations. Initially, in the first test block, there was little differentiation between different types of features. However, with more observations, there was a clear effect of prior knowledge about existing product types, such that over the course of learning, accuracy on critical features was increasingly higher than accuracy on filler features. Next, turning to performance on incongruent features, it is clear that people also learned this information well, again with better performance than filler features. Violating prior knowledge about general characteristics of cereals also led to good memory performance.

These observations were supported by a two-way repeated measures analysis of variance (ANOVA). There was a significant main effect of training block, $F(2, 90) = 101.43$, $p < .001$, $MSe = .027$, and a significant interaction between feature type and training block, $F(10, 450) = 2.99$, $p < .001$, $MSe = .059$. The main effect of feature type was also significant, $F(2, 90) = 40.38$, $p < .001$, $MSe = .053$. Although Figure 3 suggests better performance on incongruent features than critical features, a t test comparing the two types of features, pooled over blocks, did not quite reach the level of statistical significance, $t(45) = 1.90$, $p = .064$. However, the difference between filler and critical features was significant, $t(45) = 7.32$, $p < .001$.

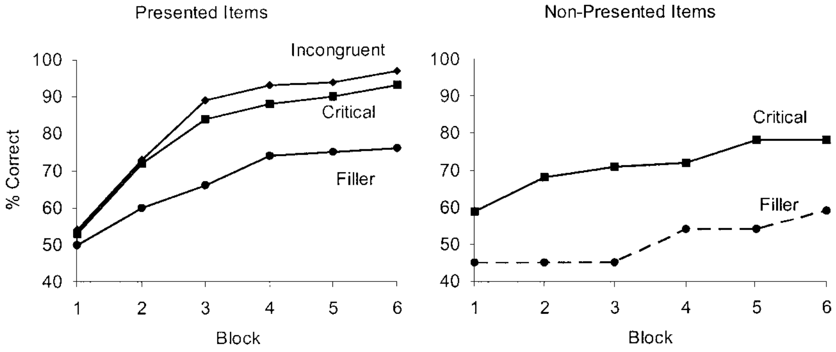

Figure 3. Results of Experiment 1.

The right side of Figure 3 shows the effect of additional study blocks on responses for nonpresented features. Performance on the nonpresented critical features is a pure measure of use of prior knowledge that is not influenced by direct observation. Performance clearly improved with more observations, again showing use of prior knowledge about adult versus child types of cereals. The increase is most notable from the first block to the second block. A one-way ANOVA indicated that the effect of study blocks on nonpresented critical features was significant, $F(5, 225) = 5.14$, $p < .001$, $MSe = .045$. The figure also shows responses on nonpresented filler features, but because these solely represent chance guessing, they were not included in the analyses.

Discussion

First, it should be noted that the main findings of Heit and Bott (2000) were replicated. Participants learned about both critical and filler features, and there was practically no advantage for critical features in the first block; however, there was an increased facilitation for critical features on subsequent blocks. In addition, there was an increasing tendency to categorize nonpresented critical features on the same basis as presented critical features. For example, even when *cartoon character on box* was never presented as a characteristic of the Daily cereal, participants would increasingly associate this with the Daily cereal when other features related to prior knowledge about children's cereals had been observed. We note that Experiment 1 tested the full set of features on each test block, whereas in the Heit and Bott study, only half the features were tested on each test block. The more frequent testing of features on consecutive blocks did not seem to have affected the results of Experiment 1.

We also note that, whereas we found a greater effect of prior knowledge after later training blocks rather than earlier training blocks, this is by no means the only possible result. Indeed, at earlier points in training, when accuracy would otherwise be close to 50% and there is the least risk of ceiling effects, there is the greatest opportunity for performance to be facilitated by prior knowledge. For example, in an unpublished experiment by Bott (2001), participants were given hints at the start of learning, indicating which prior knowledge would be most helpful to learning the categories. The effect of hints was greatest on the earliest blocks (see Heit & Bott, 2000, for other examples of knowledge having a greater benefit at earlier points in learning).

The main new result in Experiment 1 concerns incongruent features, which were learned significantly better than filler features and suggestively better than critical features, although this latter comparison did not quite reach statistical significance. According to the baseline application of the Baywatch model, in which prior knowledge would not affect incongruent features because they are associated with neither adult nor children's cereals, incongruent features would be predicted to be learned worse than critical features and at about the same level as filler features.

Hence, in light of this new result, we present two possible reconceptions of the Baywatch model. The first reconception is that incongruent information will attract extra attention. Heit (1998b) referred to this explanation as incongruent weighting and indeed found some evidence for incongruent weighting using a different experimental procedure. Although features such as *helps you get to sleep* and *flavored with bacon* would be equally nonassociated with adult and children's cereals, they would nonetheless be surprising for any kind of cereal. These incongruent features could draw more attention when they are observed, as a kind of a von Restorff effect. Therefore, even without any facilitation due to prior knowledge, additional resources could be devoted to linking incongruent features to the category labels. In terms of Figure 2, the learning rate for incongruent features would be higher than the learning rate for filler features. Thus, every time an incongruent feature is observed, more will be learned about its connection to one of the category labels, as if the incongruent feature had been presented multiple times. Note that this reconception of the Baywatch model is admittedly incomplete in some areas, for example, there is no mechanism specified for determining whether some feature is incongruent with prior knowledge. However, it will be seen that this reconception is still sufficiently developed to derive some experimental predictions.

The second reconception rests on the notion that features relating to sleeping pills and cooked breakfasts, although incongruent with prior knowledge about cereals, are actually congruent with other sources of prior knowledge, namely about sleeping pills and cooked breakfasts. Therefore, what appeared to be facilitation of incongruent features in Experiment 1 was actually facilitation of features that are congruent with sources of knowledge from other domains. The additional knowledge reconception of the Baywatch model is illustrated in Figure 4. Here, the incongruent features have associations to additional PK nodes, encompassing knowl-

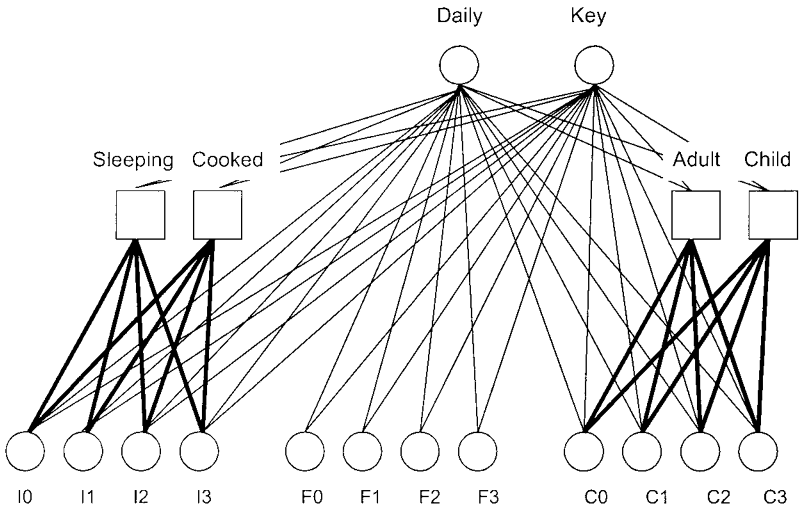

Figure 4. The Baywatch model with additional prior knowledge nodes.

edge about sleeping pills and cooked breakfasts. Just as the connections to PK nodes for adult and children's cereals would facilitate learning about critical features, the connections to PK nodes about sleeping pills and cooked breakfasts would facilitate learning about incongruent features. Note that this reconception of the Baywatch model is also incomplete, for example, more needs to be said about how these additional PK nodes are recruited or accessed. This issue, of how the Baywatch model addresses the knowledge selection problem will be considered further in the General Discussion. At this point, it is sufficient to note that the two reconceptions of the Baywatch model make differing predictions (see the Appendix for further details about implementation of simulations and illustrations of predictions).

Namely, the incongruent weighting account predicts that learning of incongruent features will be facilitated as they are observed, whereas the additional knowledge account predicts that further PK nodes are recruited. The crucial difference between these two accounts concerns categorization of features that are never presented. Say that for the breakfast cereal that induces sleep, participants are tested on the feature *prescribed by a doctor,* which had never been presented. The additional knowledge account would predict robust categorization of this feature, just as nonpresented critical features strongly followed the pattern of prior knowledge (see right panel of Figure 3). That is, presented incongruent features such as *helps you get to sleep* would lead to the recruitment of knowledge related to sleeping pills and would support the inference that this cereal is prescribed by a doctor. In comparison, the incongruent weighting account would not predict robust categorization of nonpresented incongruent features. This account does not directly provide a mechanism for drawing inferences about nonpresented incongruent features—it only addresses facilitated learning for presented incongruent features.

Experiment 2

This experiment had two aims. The primary aim was to examine learning about incongruent features that had not been presented during the study blocks but were nonetheless suggested by incongruent features that had been studied. For the cereal type that led to sleep, the nonpresented incongruent feature was *prescribed by a doctor,* and for the cooked breakfast flavored cereal, the nonpresented incongruent feature was *served in a greasy spoon café.* Note that there was no direct semantic overlap between the presented incongruent features and the nonpresented incongruent features. That is, the presented incongruent features for the sleep-inducing cereal did not mention doctors or prescriptions, and the presented incongruent features for the cooked breakfast cereal did not mention any kind of restaurant or café.

If the facilitation of incongruent features, exhibited in Experiment 1, was due to recruiting additional sources of prior knowledge, for example, about sleeping pills and cooked breakfasts, then this prior knowledge should also facilitate classification about nonpresented incongruent features that are related to this prior knowledge. In contrast, if the facilitated learning of incongruent features was due to greater attention during study, as in the incongruent weighting account, there would be no facilitated learning of incongruent features that were not directly studied. The only way for a presented incongruent feature such as *helps you get to sleep* to facilitate classification of a nonpresented incongruent feature such as *prescribed by a doctor* would be through mediation by other knowledge of sleeping pills, medicines, and so on. The mere surprisingness of *helps you fall asleep* as a breakfast cereal feature would not predict strong performance on *prescribed by a doctor.*

The secondary aim was to see how the presence of incongruent features affected learning of other features, namely critical and filler features. Therefore, the condition with incongruent features was compared with a control condition without incongruent features. In general terms, the Baywatch model operates on the basis of cue competition: Different features compete to predict the category labels. If the relations between incongruent features and category labels are learned well, then there should be some decrement in learning about critical features. In the control condition, the incongruent features were replaced with features concerning where the cereal was tasted, either at home or in a shop. These control features made the incongruent and control conditions comparable in terms of number of features presented. In terms of content, the control features were like additional filler features in that they were intended to be generic information that could be compatible with any breakfast cereal. Hence, it was predicted that the control features, concerning where the cereal was tasted, would not be learned as well as the incongruent features, concerning sleeping pills and cooked breakfasts. Furthermore, due to competition, the critical features in the incongruent condition should be learned somewhat worse than the critical features in the control condition.

Method

Experiment 2 was like Experiment 1, with the following changes. There were 80 participants, 40 in the incongruent condition and 40 in the control condition. The incongruent condition was exactly like Experiment 1 except that in the test blocks, two additional, nonpresented incongruent features were tested: *prescribed by a doctor* and *served in a greasy spoon café.* Hence, the test lists had 58 items. The nonpresented incongruent features were selected on the basis of a pretest in which 14 participants rated the strength of association between possible features and the themes of cooked breakfasts and inducing sleep. These two nonpresented features were rated as most highly associated to their respective themes. In the control condition, incongruent features were replaced with four variants relating to the theme of where the cereal was tasted, that is, where the market research interview had been conducted. For cereals tasted at home, the presented features were *tasted at detached house, tasted at flat,* and *tasted at terraced house,* and the nonpresented feature was *interviewed at own home.* For cereals tasted at a retail outlet, the presented features were *tasted at supermarket, tasted at grocery store,* and *tasted at shop,* and the nonpresented feature was *interviewed at retail outlet.*

Results

Before reporting the statistical evidence, we summarize the key findings, as shown in Figure 5. The main comparisons of interest were in the incongruent condition. In the top two panels of this figure it is clear that for both presented and nonpresented features, incongruent features were learned better than critical features, which were in turn learned better than filler features. In other respects, the incongruent condition replicates the results of Experiment 1. Of secondary interest was the comparison between the incongruent condition and the control condition, shown in the bottom two panels. Critical features were learned somewhat better in the control condition than in the incongruent condition, although this difference is small, particularly for presented features.

Presented features. The first statistical analysis was a three-way ANOVA for the presented features, with experimental condition, feature type, and training block as variables. Note that in terms of the experimental design, the control features replaced the incongruent features; hence, the feature types in the ANOVA were incongruent, critical, and filler in the incongruent condition and control, critical, and filler in the control condition. There was a main effect of experimental condition, $F(1, 78) = 9.10$, $p < .001$, $MSe = .182$, and a significant interaction between experimental condition and feature type, $F(2, 78) = 65.20$, $p < .001$, $MSe =$

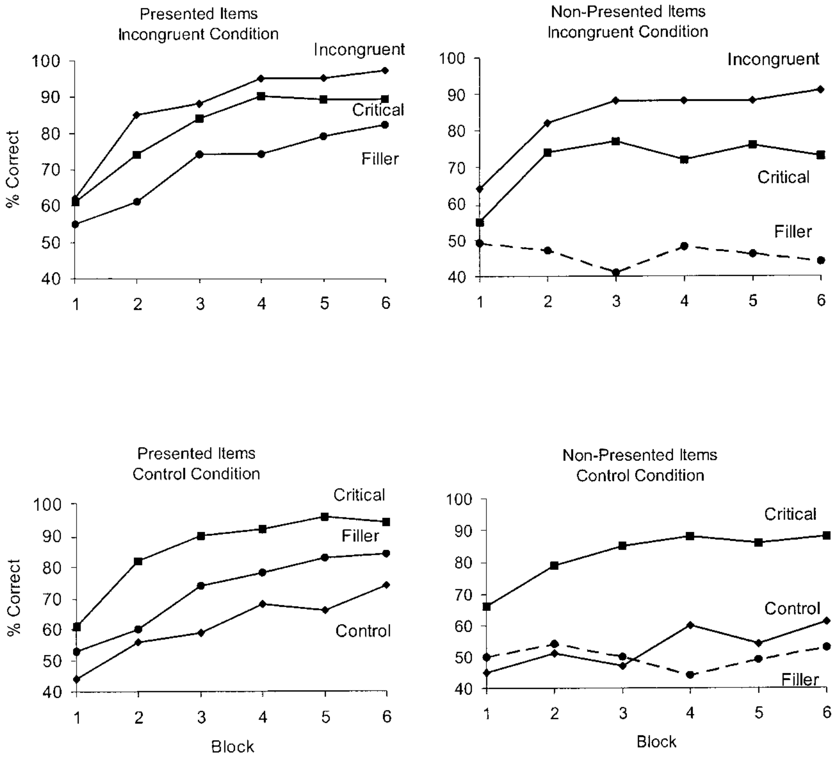

Figure 5. Results of Experiment 2.

.005. The interaction between experimental condition and training block was not significant, *F*(5, 390) $<$ 1. There were significant main effects of training block, *F*(5, 390) = 107.32, $p < .001$, *MSe* = .003, and feature type, *F*(2, 156) = 39.65, $p < .001$, *MSe* = .005. The interaction between training block and feature type was also significant, *F*(10, 780) = 2.71, $p < .01$, *MSe* = .002. Finally, the three-way interaction did not reach the level of statistical significance, *F*(10, 780) = 1.66, *MSe* = .002.

Just focusing on the incongruent condition, we compared overall performance on the three types of features. Using *t* tests with Bonferroni corrections applied, all three pairwise comparisons (incongruent vs. critical, incongruent vs. filler, and critical vs. filler) were statistically significant at $p < .01$. Hence, as suggested by Figure 5, learning was significantly better for incongruent features than for critical features and significantly better for critical features than for filler features.

A further ANOVA was used to compare just the critical and filler features in the two conditions, allowing an assessment of how the presentation of incongruent features versus control features affected learning of the other features. There was no significant main effect of experimental condition, *F*(1, 78) = 1.21, *MSe* = .152, and no significant interaction between experimental condition and feature type, *F*(2, 78) = 2.37, *MSe* =.004. Hence, there did not seem to be a significant effect of presenting incongruent versus control features on the learning of presented critical and filler features. The interaction between experimental condition and training block was not significant, *F*(5, 390) = 1.01, *MSe* = .002. There were significant main effects of training block, *F*(5, 390) = 111.75, $p < .001$, *MSe* = .002, and feature type, *F*(1, 78) = 101.66, $p < .001$, *MSe* = .004. The interaction between training block and feature type was also significant, *F*(5, 390) = 3.76, $p < .01$, *MSe* = .002. Finally, the three-way interaction did not reach the level of statistical significance, $F < 1$.

Nonpresented features. We next turn to the results for nonpresented critical and incongruent features. As in Experiment 1, filler features were not analyzed. A three-way ANOVA showed a significant main effect of experimental condition, *F*(1, 78) = 9.23, $p < .001$, *MSe* = .249. There was a significant interaction between experimental condition and feature type, *F*(2, 117) = 52.22, $p < .001$, *MSe* = .192, but the interaction between experimental condition and training block was not significant, *F*(5, 390) = 1.37, *MSe* = .006. There was a main effect of feature type, *F*(1, 78) = 8.70, $p < .01$, *MSe* = .192. The main effect of training block was significant, *F*(5, 390) = 15.67, $p < .001$, *MSe* = .006. The interaction between feature type and training block was not significant, $F < 1$, and the three-way interaction did not reach the level of significance, $F < 1$.

Next, just focusing on the incongruent condition, we compared overall performance on critical versus incongruent features. Using a paired *t* test on the data pooled over blocks, performance was significantly better for incongruent features, *t*(39) = 3.46, $p < .01$.

Finally, overall performance on nonpresented critical features was compared between the two conditions, and this difference was found to be statistically significant, *t*(78) = 2.52, $p < .05$. That is, presenting incongruent features rather than control features did lead to a significant decrement in performance on classifying critical features based on prior knowledge.

Discussion

The most important result from Experiment 2 was in the incongruent condition in which performance on incongruent features was significantly better than performance on critical features for both presented features and nonpresented features. It is clear that the incongruent features were learned very well despite their lack of predictability on the basis of prior knowledge about breakfast cereals.

The results for nonpresented features are particularly informative. According to the additional knowledge reconception of the Baywatch model, prior knowledge about sleeping pills and cooked breakfasts would be recruited. This account predicts that nonpresented incongruent features will also be classified robustly according to prior knowledge, and the results supported this account. In contrast, the incongruent weighting reconception of the Baywatch model has a means for predicting robust classification of presented incongruent features but does not directly predict this level of classification for nonpresented incongruent features. Hence the results on nonpresented incongruent features favored the additional knowledge account over the incongruent weighting account. That is, it appears that participants recruited knowledge from outside the domain of breakfast cereals to facilitate learning and classification of incongruent features, whether or not they were observed.

Looking at the finer details of Figure 5, we note that nonpresented incongruent features were learned even better than nonpresented critical features. In terms of the additional knowledge version of the Baywatch model, this finding could be explained in terms of stronger prior knowledge about sleeping pills and cooked breakfasts than about children’s and adults’ cereals (see Appendix for implementation details). That is, in the model, PK nodes could facilitate responses on both incongruent and critical features, but the PK nodes corresponding to incongruent features might correspond to stronger beliefs or better established categories. In terms of Figure 4, the PK nodes for sleeping pills and cooked breakfasts could have stronger outputs than the PK nodes for children’s and adults’ cereals. The additional knowledge account does not make a firm prediction about whether critical or incongruent features will be learned better. Instead, its key prediction, and what distinguishes it from incongruent weighting, is that there will be some advantage for nonpresented incongruent features over nonpresented filler features.

We next consider the results of the presented features in the incongruent condition, where the pattern again was incongruent features better than critical features which were better than filler features. In terms of the Baywatch model, there are at least two ways to explain this pattern. The first is with the additional knowledge reconception of the model, by assuming that prior knowledge relating to the incongruent features was stronger than prior knowledge relating to the critical features. This is the same assumption that would be made to explain the pattern for nonpresented features, but again, note that this account does not make a firm prediction here. The second possibility is in terms of incongruent weighting. Depending on the learning rate for incongruent features, it is possible for performance on presented incongruent features to surpass that of presented critical features. Performance on critical features would already be boosted because of PK nodes, but if incongruent weighting is particularly strong, there may be even a greater boost for incongruent features. This too is not a firm prediction of the incongruent weighting account, but is parameter dependent (see the Appendix for implementation details).

In sum, all of the results from the incongruent condition can be explained in terms of additional PK nodes, corresponding to relatively strong beliefs about sleeping pills and cooked breakfasts, facilitating performance on incongruent features. However, we do not rule out any possible role for incongruent weighting (which Heit, 1998b, did report). Although incongruent weighting cannot explain the results for nonpresented incongruent features, it is possible that the strong performance on presented incongruent features could be in part because of incongruent weighting.

It is also useful to compare the control condition to the incongruent condition. The strongest result was that control features, relating to the interview being conducted at home or in a shop, were not learned well compared with either incongruent features or critical features. Performance on control features was even somewhat worse than performance on filler features, although the Baywatch model does not make a prediction for this comparison. Whereas there could be some small amount of prior knowledge recruited concerning homes and shops that would be relevant to these control features, it seemed that this prior knowledge did not correspond to discrete, coherent, and separable categories of consumer products in the same way as adults' versus children's cereals and sleeping pills versus cooked breakfasts. That is, most foods could be tasted at home or out of the home, just as most cereals could have either a white box or a cream-colored box. It was predicted that because of the competitive nature of feature learning in the Baywatch model, in the incongruent condition the robust learning of incongruent features would have some detrimental effect on learning about critical features compared with learning about critical features in the control condition. Comparison of the two conditions in Figure 5 suggests that presented critical features were learned slightly worse in the incongruent condition than in the control condition, but this difference was not statistically significant. However, in terms of a more sensitive measure, judgments on nonpresented critical features, there was a significant detrimental effect in the incongruent condition compared with the control condition. It might be possible to derive other predictions from the Baywatch model regarding cue competition; however, overall there did not seem to be strong evidence for this aspect of the model (see also Kaplan and Murphy, 2000, concerning possible competition between learned features due to use of prior knowledge).

Experiment 3

Experiments 1 and 2 showed robust learning of incongruent features. However, these incongruent features, from outside the domain of breakfast cereals, are by no means the only possible kind of incongruent feature. There are many possible ways to be incongruent with prior knowledge. Experiment 3 created incongruent features that were contradictory to prior knowledge by presenting some features typical of adult cereals within product descriptions for the child cereal type and vice versa. For example, the Daily cereal might generally have adult features such as *low in sugar* and *health tips on box* but also have a smaller number of child features such as *cartoon character on box.*

In related work, Kaplan and Murphy (2000) compared category learning with mixed-theme features to category learning with intact-theme features. In the intact theme condition, participants learned about a pair of categories with features that were consistent with prior knowledge, for example, features of one category were related to jungle vehicles and features of the other category were related to arctic vehicles. In the mixed theme condition, the knowledge-related features were mixed up so that a category might have 50% of the jungle features and 50% of the arctic features. Kaplan and Murphy reported that learning was worse in the mixed theme condition, suggesting that there was more facilitation due to prior knowledge in the intact theme condition (see also Murphy and Kaplan, 2000).

Our own Experiment 3 made a similar comparison but within a single condition in which 67% of the knowledge-related features in a category, the critical features, fit a single theme, and 33% of the knowledge-related features, the contradictory features, fit the theme of the contrasting category. By using the same methodology as Experiments 1 and 2, we were able to collect detailed information about the time course of learning critical, contradictory, and filler features.

Experiment 3 served as a further test of the Baywatch model. Although this model can be modified to account for facilitated learning of incongruent features coming from an outside domain, even this modified model would not predict facilitation on contradictory features. The contradictory features would not be able to recruit prior knowledge in the way that incongruent features about sleeping pills and cooked breakfasts would; thus, there would be no facilitation for contradictory features. Although the contradictory features would still tend to activate the PK nodes for children's and adults' breakfast cereals, in general these PK nodes would be poor predictors of the category labels, Daily and Key. Indeed, presentation of contradictory features would directly disrupt learning of the links between these PK nodes and the category labels. Hence, the model would predict that the influence of prior knowledge would be much weaker overall in this experiment compared with Experiments 1 and 2. In terms of Figure 2, participants were predicted to rely mainly on direct empirical learning of associations between input features and output category labels, without mediation of the PK nodes (see the Appendix for further details of this simulation). To the extent that prior knowledge would have some small effect, the model would predict that critical features would be facilitated relative to filler features and that learning on contradictory features would be worse than filler features. Therefore, it was also predicted that, in contrast to incongruent features in Experiments 1 and 2, there would be no facilitation for contradictory features.

Method

Experiment 3 was like the control condition of Experiment 2 except for the following. The eight pairs of critical features were assigned, randomly for each participant, as follows: Four pairs served as presented critical features, two pairs served as nonpresented critical features, and the remaining two pairs were contradictory features, that is, assigned to the opposite category. For example, the Daily cereal might have four presented adult features, two nonpresented adult features, and two presented child features. Likewise, the Key cereal would have four presented child features, two nonpresented child features, and two presented adult features.

Each description had two filler features as well as two other features that might be both critical, both contradictory, or one critical and one contradictory. The contradictory features were randomly assigned to individual product descriptions in the study blocks and could appear either as two contradictory features within a single description or as one contradictory feature in each of two product descriptions. Thus, over the three descriptions presented for each cereal type, there were four presented critical features that were typical of the product to which they were allocated and two contradictory features that were typical of the other product type. As in the control condition of Experiment 2, each description also had one control feature, but these were not subject to any hypotheses of interest.

There were 40 participants.

Results

Mean accuracy results for critical, contradictory, and filler features are shown in Figure 6. The results for presented features show that performance with contradictory features was poorer than for critical and filler features throughout most of the six practice blocks. Performance with critical features was also somewhat poorer than in the previous experiments but still surpassed performance on filler features. A two-way repeated measures ANOVA on the features presented during learning showed that the main effect of feature type did not quite reach significance, $F(2, 78) = 2.69$, $p = .074$, $MSe = .009$. The main effect of block was significant, $F(5, 39) = 32.09$, $p < .001$, $MSe = .004$, as was the interaction between feature type and block, $F(10, 390) = 3.38$, $p < .001$, $MSe = .004$. This interaction was not anticipated, but it was examined with paired *t* tests at each level of the block variable, comparing the three feature types to each other with Bonferroni corrections applied. Only one of these comparisons reached the level of statistical significance, namely that on the third test block, performance was significantly greater on critical features than on contradictory features, $p < .05$.

A one-way repeated measures ANOVA on the nonpresented critical features showed that the effect of block was significant, $F(5, 195) = 3.60$, $p < .01$, $MSe = .005$; hence, there did seem to be some change in use of prior knowledge with more observations.

Discussion

The main result was that learning of contradictory features was not facilitated, as was learning of incongruent features in Experiments 1 and 2. We note that previous reviews of prior knowledge effects on categorization (e.g., Heit, 2001a; Murphy, 2002) have not emphasized this distinction between two kinds of incongruent features. As predicted by the Baywatch model, simply violating prior knowledge does not always lead to strong learning. Indeed, there was not a significant main effect of feature when comparing critical features that fit prior knowledge, contradictory features that went against prior knowledge, and filler features that were neutral with respect to prior knowledge. (Bott, 2001, who used a similar design, but with building stimuli rather than breakfast cereal stimuli, also found no main effect of feature type when contradictory features were present.) As in Experiments 1 and 2, there was a significant interaction between feature type and block, but we believe that the pattern of interaction on the left side of Figure 6 should not be overinterpreted. The first block included a high degree of chance responding, and in the last block there were virtually no differences between features, as if all feature types were learned mainly by rote.

Overall, it does seem that, as predicted, prior knowledge had little influence on learning. In contrast to Experiments 1 and 2, there was no main effect of feature type on learning of presented features. The nonpresented critical features would be the most sensitive measure of prior knowledge use (see right side of Figure 6). Although there was a significant effect of block, indicating some use of prior knowledge, the proportion of prior knowledge-based classifications in the last block was only 66%, which was lower than in the previous two experiments. In terms of the Baywatch model, because the PK nodes would not be much help in predicting category labels, learning mainly took place using the empirical part of the network, relying on direct connections between input nodes and category labels.

We note finally that the lack of facilitation on contradictory features does not lend much support for the incongruent weighting account, which would seem to predict enhanced learning for these features which, by design, went against participants' expectations.

General Discussion

These three experiments support the general idea that observations of category members are used to recruit prior knowledge. This point is made most directly by performance on the critical features. As more category members were observed, the critical features were increasingly classified in accordance with prior knowledge. Even the critical features that had never been presented showed this pattern. In Experiments 1 and 2, a similar

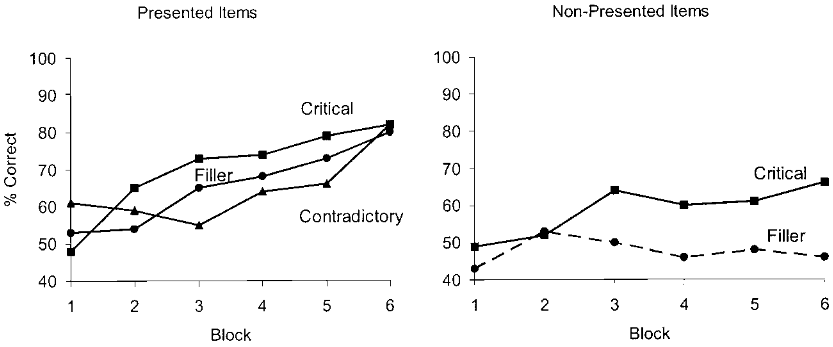

Figure 6. Results of Experiment 3.

explanation can be given for the strong performance on incongruent features (including nonpresented incongruent features in Experiment 2). It appears that people recruited additional knowledge from outside the domain of breakfast cereals to facilitate responses on incongruent features. Although the idea of extra attentional weight to incongruent information cannot explain the pattern of results for nonpresented features, it is possible that incongruent weighting had some contribution to the results for presented features. Finally, in Experiment 3, participants observed a different kind of incongruent feature, which contradicted prior knowledge. Because prior knowledge, even from outside the domain of breakfast cereals, would not support learning of these features, there was no facilitation of contradictory features. Indeed, there appeared to be a reduced use overall of prior knowledge in Experiment 3 as if the presence of the contradictory features encouraged rote learning.

The original application of the Baywatch model (Heit & Bott, 2000) did not address learning of any kind of incongruent features. In light of the present results, the modified version of the Baywatch model in Figure 4 does seem to be a useful way to describe the time course of applying prior knowledge to learning critical, filler, and incongruent features. However, a salient potential criticism of the Baywatch model is that although it may describe the results, it depends on selecting useful sources of prior knowledge from an extremely large database. For example, if the model can use prior knowledge about breakfast cereals, cooked foods, and medicines, it must be assumed that there is access to a very wide range of knowledge sources. How does the model select useful prior knowledge from all these potential sources and actually use it to make learning easier? Put another way, Figure 4 only shows four PK nodes, but a more realistic description would have to show a much larger number of PK nodes. In fact, the Baywatch model was originally intended to address this problem of knowledge selection; hence, in the following section we discuss this issue.

Knowledge Selection

Heit and Bott (2000) compared the issue of knowledge selection in category learning with the issue of hypothesis selection in Bayesian statistical estimation (e.g., Raiffa & Schlaifer, 1961), which provides many techniques for combining multiple prior beliefs with observations and selecting from these beliefs based on the data observed. In Bayesian statistics, there is no assumption that a learner starts with optimal or perfectly correct prior beliefs. Instead, the learner begins with reasonable guesses that merely serve as an initial basis for learning, with corrective information then provided by the data. Indeed, it is possible to start with numerous different prior beliefs, with a distribution of initial degrees of confidence in each of these. When observations are made, confidence in various prior beliefs can be increased or decreased as appropriate. That is, observations can be used to select from different sources of prior knowledge. Still, it might be argued that even Bayesian statistics does not fully address the knowledge selection problem, because these methods merely indicate how to select from a set of prior beliefs, but they do not say which prior hypotheses should be in the starting set. The key point is that Bayesian techniques can be applied to a large set of prior beliefs even when many of them are repetitive or poorly chosen, as long as this set covers the hypothesis space well enough so that the target concept can be represented.

Turning more specifically to the present experiments, it seems that there might be two problems in applying the Baywatch model to the use of knowledge in category learning. Namely, the many possible sources of prior knowledge will include some that are repetitive or even poor predictors of the categories to be learned. First, would there be a knowledge selection problem if there are repetitive sources of prior knowledge? For example, features of the Daily cereal might be predicted on the basis of prior knowledge of adult cereals or prior knowledge of healthy cereals. Likewise, some people might distinguish between sleeping pills and herbal sleep remedies, and either source of prior knowledge would be useful for predicting incongruent features. Would this redundancy in prior knowledge somehow worsen the knowledge selection problem or otherwise degrade performance of the Baywatch model? Heit and Bott (2000) investigated this question by conducting additional simulations using the example of learning to distinguish church-like buildings from office-like buildings. Additional PK nodes, similar to existing PK nodes, were added to the network in Figure 1. For example, PK nodes corresponding to cathedrals and industrial parks were added, embodying much of the same prior knowledge as the PK nodes for churches and office buildings. This redundancy did not lead to problems for the model. Indeed, to the extent that redundant sources of prior knowledge were mutually supporting, having multiple sources of prior knowledge helped performance, for example, by enhancing the advantage of critical features over filler features.

A second potential problem of knowledge selection is that there will be many possible sources of knowledge that are poor choices. In the present experiments, prior knowledge about cereals, cooked foods, and medicines would all be helpful for predicting the features of the categories to be learned. In comparison, prior knowledge about coffee, exercise equipment, and digital cameras would be poor choices, and presumably there would be many more possible poor choices of prior knowledge. How does the Baywatch model select the more useful choices and ignore the poor choices? Referring to Figure 4, it is possible to imagine that many irrelevant PK nodes could be added on the left side of the diagram. Knowledge selection takes place at two locations in this network. One location is at the connections between input nodes and the PK nodes. If irrelevant PK nodes do not have any overlap with the input, then these irrelevant PK nodes will not be activated. For example, in the present experiments, the training stimuli had practically no features that would overlap with knowledge of exercise equipment or digital cameras; thus, these sources of prior knowledge were not activated. Adding PK nodes that are never activated would have no effect at all on the Baywatch model. The other location is at the connections between PK nodes and category label nodes. Here, different sources of prior knowledge compete to predict the category labels. Say, for example, that in the present experiments, the training stimuli had some featural overlap with prior knowledge about coffee. Hence, a PK node for coffee might be activated sometimes but not as often as PK nodes for adults' and children's breakfast cereals; thus, learning the connection between coffee and the category labels would be slower. In fact, prior knowledge about coffee would be a poor discriminator between the Daily and Key categories in these experiments, and as a result hardly anything would be learned at all about links between the PK node for coffee and the category labels. Again, adding PK nodes that are sometimes activated but are poor predictors of the category labels would have little effect on performance of the Baywatch model (see also Heit & Bott, 2000, for information about these simulations as well as a discussion of "malicious" PK nodes, which might intermix beliefs from two other PK nodes, such as a PK node that would have some sleeping pill features and some cooked breakfast features).

In sum, the Baywatch model gives a suggestive account of ways that many sources of prior knowledge could be used—or not used—in category learning. To the extent that multiple sources are repetitive, there is a facilitative effect and not a serious knowledge selection problem. Selecting the useful sources and ignoring the poor sources would take advantage of the content of the prior knowledge and the competitive nature of learning. That is, irrelevant sources of prior knowledge would have little overlap with the input stimuli, and even if there is some overlap, these sources would be poor competition for more appropriate sources of prior knowledge.

Alternative Accounts

Although most modeling work in categorization research has not addressed prior knowledge effects, the methods used to incorporate prior knowledge into category learning by the Baywatch model are by no means the only possible way to address this issue (see Heit & Bott, 2000, for a review). Heit (1994, Heit 1998b, 2001b) assessed variations of the integration model of prior knowledge effects on categorization. That model was in some ways a precursor of the Baywatch model, in which the connections between prior knowledge and the output categories are given to the model in advance. That is, the model does not learn to select from different sources of knowledge. Hence, the integration model could not explain the present experiments investigating the knowledge selection issue.

Recently, Rehder and Murphy (2003) presented a psychological model called KRES (for knowledge resonance) and applied it successfully to several experiments on prior knowledge effects on category learning, including the results from Heit and Bott (2000). KRES has a different architecture and learning mechanism than Baywatch: It is a recurrent network with Hebbian learning. However, the most crucial difference is that KRES does not rely on PK nodes corresponding to existing categories, although it can use them, and allows prior knowledge to be represented flexibly in terms of a set of interconnections within a feature set. Rehder and Murphy suggested that its reliance on PK nodes might limit the applicability of the Baywatch model to learning about categories that are like known categories; however, this suggestion seems to be contradicted by the present research. Here, the Baywatch model has been applied to a situation in which people learned about unfamiliar categories, for example, breakfast cereals that put you to sleep, by combining prior knowledge from two existing categories, breakfast cereals and sleeping pills. Therefore, Baywatch is not limited to learning about known categories.

Note that the present experiments were not intended to distinguish Baywatch from KRES. In some ways, KRES can be thought of as a superset of Baywatch. KRES represents knowledge in such a flexible manner—in terms of any possible connection between features—it seems likely that any set of beliefs represented by Baywatch could also be represented by KRES. To distinguish KRES from Baywatch could well require focusing on predictions derived from the distinctive aspects of KRES, namely its recurrent architecture and Hebbian learning rule. For example, Rehder and Murphy (in press) suggested that when the evidence for feature competition is weak or mixed, as in our own Experiment 3 and in Kaplan and Murphy (2000), KRES may give a better account than other models, because it does not necessarily predict competition. We believe strongly that the wider gap between models, and the more important distinction, is between categorization models that do incorporate prior knowledge and those that do not. The purpose of the present experiments was to test and refine the Baywatch model by applying new data concerning incongruent features. It seems plausible that versions of the KRES model could also be applied to these data, and likewise this process would be informative for the development of KRES. One important assumption that current applications of the Baywatch model share with KRES is local, atomistic representation of input features. That is, each input feature, such as *high in sugar,* is represented as a single input node. Local representation of input features is a frequent assumption in connectionist network models (e.g., Gluck & Bower, 1998), and the aim of our work has been to show how knowledge selection can be addressed in a model with local representation of inputs. Another possibility is that input features could have distributed representations over a vector of microfeatures. For example, *high in sugar* could be represented as a particular pattern in the input vector, and *animal-shaped* could be represented as another pattern. Some of the information encoded in PK nodes, namely real-world co-occurrence, could instead be encoded in the pattern of distributed input representations (see Landauer & Dumais, 1997). Hence, the overlap in these two patterns could reflect that these two features do co-occur in some contexts such as children's cereals.

Conclusion

The issue of how multiple sources of prior knowledge are used along with observations is one of great generality for cognitive psychology and cognitive science. The account presented here fits with Sperber's (1994) conception of *partial modularity.* There are clear benefits to applying prior knowledge to categorization, and there may well be benefits to representing knowledge as encapsulated modules corresponding to different domains or different subtypes within a domain. However, to apply knowledge in a productive way, it is necessary to cross domain boundaries and integrate knowledge from multiple sources. One attractive way of doing so is to allow perceptual inputs, or verbal descriptions of inputs, to serve as a gate for selecting which knowledge is used. Hence, descriptions of bran flakes that put people to sleep would access knowledge about breakfast cereals and sleeping pills (for additional examples of the use of gating of inputs to select from different sources of prior knowledge, see Hayes et al., 2003; Lewandowsky et al., 2000; and Yang & Lewandowsky, 2003). The present research shows that observing descriptions of category members can help to recruit multiple sources of knowledge; thus, seemingly incongruent features can become predictable on the basis of knowledge from other domains and serve as a reliable element of category representation.

References

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology, 36,* 203–272.

Bott, L. (2001). *Prior knowledge and statistical models of categorization.* Unpublished doctoral thesis, University of Warwick, United Kingdom.

Bott, L., & Heit, E. (2004). Nonmonotonic extrapolation in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 28–37.

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

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

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

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

Heit, E. (1998a). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248–274). Oxford, England: Oxford University Press.

Heit, E. (1998b). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 712–731.

Heit, E. (2001a). Background knowledge and models of categorization. In U. Hahn & M. Ramscar (Eds.), *Similarity and categorization* (pp. 155–178). Oxford, England: Oxford University Press.

Heit, E. (2001b). Putting together prior knowledge, verbal arguments, and observations in category learning. *Memory & Cognition, 29,* 828–837.

Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D. L. Medin (Ed.), *Psychology of learning and motivation* (Vol. 39, pp. 163–199). San Diego, CA: Academic Press.

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

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

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Lewandowsky, S., Kalish, M., & Griffiths, T. L. (2000). Competing strategies in categorization: Expediency and resistance to knowledge restructuring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1666–1684.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69,* 397–407.

Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Murphy, G. L., & Kaplan, A. S. (2000). Feature distribution and background knowledge in category learning. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 53A,* 962–982.

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

Palmeri, T. J., & Blalock, C. (2000). The role of background knowledge in speeded perceptual categorization. *Cognition, 77,* B45–B57.

Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory.* Boston: Harvard University, Graduate School of Business Administration.

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

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38,* 495–553.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp.39–67). Cambridge, MA: Cambridge University Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–641.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–282.

Yang, L.-X., & Lewandowsky, S. (2003). Context-gated knowledge partitioning in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29, 838–849.

Appendix

Simulations of the Baywatch Model

Predictions of Baseline Model for Experiment 1

The first simulation made similar assumptions to Heit and Bott (2000). The simulation used a reduced set of training examples compared with the actual Experiment 1, because in effect the three Daily training examples and the three Key training examples had the same structure but applied to different features. In the simulation, there was just one Daily training example with two critical features presented for this category, and likewise there was one Key training example with two other critical features presented.

The structure of the network is illustrated in Figure 2 (except that there were six filler input nodes and six critical input nodes). Each input feature was assigned a different input node in the network, with a 1 value coding presence of the feature and a 0 value coding absence. The two output units varied continuously between -1 and $+1$. One output unit corresponded to each category. The activation on a category was given by the weighted sum of its inputs. This activation was then converted into a probability measure using the logistic transformation given in Gluck and Bower (1988, Equation 7). If a Daily example was presented during training, the teaching values for the category nodes would be $+1$ on the Daily output node and -1 on the Key output node. These values would be reversed for a Key training example.

(*Appendix continues*)

Inc	Inc	Inc	Inc	Fill	Fill	Fill	Fill	Fill	Fill	Crit	Crit	Crit	Crit	Crit	Crit	Daily	Key
Pres	Non	Pres	Non	Pres	Pres	Non	Pres	Pres	Non	Pres	Pres	Non	Pres	Pres	Non	Out	Out
Daily	Daily	Key	Key	Daily	Daily	Daily	Key	Key	Key	Daily	Daily	Daily	Key	Key	Key		
Daily Training Example																	
1	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	+1	-1
Key Training Example																	
0	0	1	0	0	0	0	1	1	0	0	0	0	1	1	0	-1	+1

Figure A1. Training and test stimuli for simulations of Experiments 1 and 2. Inc = incongruent; Fill = filler; Crit = critical; Pres = presented; Non = nonpresented; Out = output.

The critical features were connected by fixed weights to the prior knowledge nodes, with values of either +1 or −1, corresponding to prior knowledge of whether features are positively or negatively associated with the PK nodes. For example, the critical feature *high in sugar* had a positive connection to the PK node for children's cereals and a negative connection to the PK node for adult cereals. The output of a PK node was a threshold transformation of the weighted sum of its inputs, such that the output was 1 if the sum was greater than or equal to 1 and was 0 otherwise.

For the simulations of Experiments 1 and 2, two variants were run. In one variant, the incongruent input nodes were connected to the PK nodes for children's and adult cereals with fixed weights of −1. This variant captures the notion that all of the incongruent features would be unexpected for both kinds of cereals. However, the negative weights did not affect the results, because the PK nodes were activated by threshold functions. Each training example had two critical features and one incongruent feature; hence, the net input to a PK node was +1, which was sufficient to meet the threshold. In another variant, the connections between incongruent input nodes and these PK nodes were or fixed at a weight of 0. The same predictions were obtained from this variant.

All of the nonfixed weights in the network were initially zero and were adjusted according to the standard delta rule (e.g., Gluck and Bower, 1988). The network was trained for nine epochs, with the learning rate in the delta rule set at 0.08 and the probability mapping constant for the logistic transformation function set at 7.0. The two training examples for Experiment 1 are shown in Figure A1, one Daily description and one Key description. Note that the nonpresented incongruent features were not used in simulations for Experiment 1. The nonpresented filler and critical features were always trained with an input value of 0.

Following each training epoch, the network was tested on the individual features by presenting a vector of all zeroes except for the tested feature, which had a value of 1. The predictions are displayed in Figure A2, with the proportion correct on the test set shown as a function of the number of learning epochs and feature type. The left panel shows predictions for presented features. The model predicts that critical features will be learned faster than filler features, with an increasing advantage over the course of learning. The model predicts the same performance for incongruent features as filler features. For completeness, the right panel shows the predictions for nonpresented features, that prior knowledge will lead to increasingly better performance on critical features.

Predictions of Model With Incongruent Weighting for Experiment 2

To simulate Experiment 2, the same set of training and test items shown in Figure A1 was used, with the nonpresented incongruent features now included. The first simulation for Experiment 2 embodied the idea of incongruent weighting, namely that incongruent features would have a faster learning rate as if they had been presented more frequently or more intensely than other features. This simulation used a network structured as in Figure 2. The only change to the network was that the learning rate for

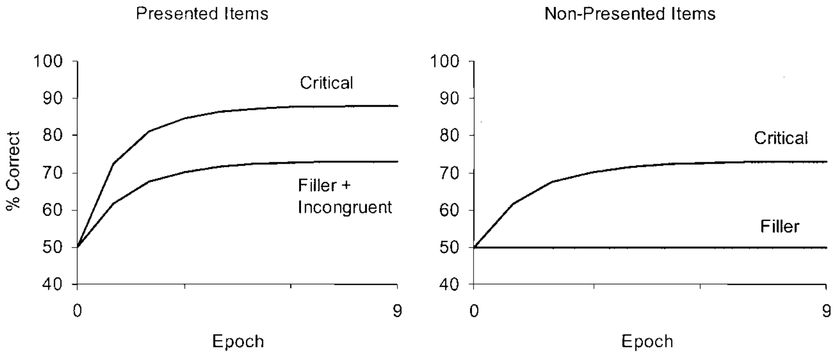

Figure A2. Predictions of the baseline model for Experiment 1.

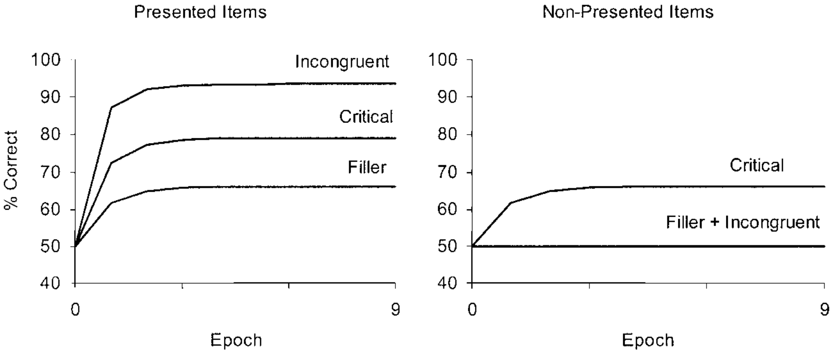

Figure A3. Predictions of the model with incongruent weighting for Experiment 2.

incongruent features was a multiple of the learning rate for filler and critical features (0.08). It was found by simulation that the value of this multiple had a great impact on the predictions for presented features. If this multiple is greater than 1 and less than 2, the model predicts that incongruent features will be learned better than filler features but worse than critical features. If the multiple is 2, the model predicts that incongruent features will be learned better than filler features and at the same level as critical features. If the multiple is greater than 2, the model predicts that incongruent features will be learned better than both critical and filler features.

The predictions for incongruent weighting are shown in Figure A3 for a multiple of 4. The left panel shows predictions for presented features. The model predicts that critical features will be learned faster than filler features and that incongruent features will be learned faster than critical features. For a multiple of exactly 2, the qualitative prediction is that the incongruent features would be at the same level as the critical features. For a multiple between 1 and 2, the prediction is that incongruent features would fall below critical features and above filler features. The right panel shows predictions for nonpresented features for a multiple of 4. The model predicts an advantage for critical features over filler features but does not predict any learning for incongruent features for any value of the multiple.

Predictions of Model With Additional Prior Knowledge Nodes for Experiment 2

The second simulation for Experiment 2 embodied the idea that additional prior knowledge nodes would be recruited corresponding to prior knowledge about sleeping pills and cooked breakfasts. Hence the network was structured as in Figure 4. The incongruent features were connected by fixed weights to these additional PK nodes, with values of either +1 or −1, corresponding to prior knowledge of whether features are positively or negatively associated with the PK nodes. For example, the incongruent feature *helps you get to sleep* had a positive connection to the PK node for sleeping pills and a negative connection to the PK node for cooked breakfasts. The learning rate for all features was 0.08.

The predictions are shown in Figure A4. The left panel shows predictions for presented features. The model predicts that critical features will be learned faster than filler features, and the model does not predict a difference between incongruent features and critical features. However, in a similar manner to the predictions of the incongruent weighting simulations, these predictions for incongruent versus critical features are parameter-dependent. When each PK node has an output value of 1, the predictions are as in Figure A4. However, it would be possible for different PK nodes

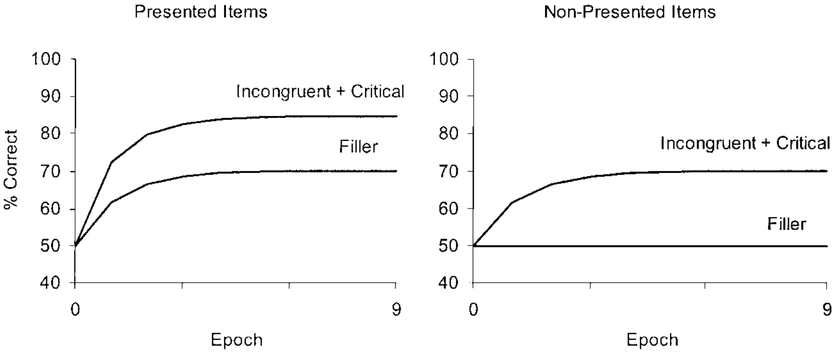

Figure A4. Predictions of the model with additional prior knowledge nodes for Experiment 2.

(*Appendix continues*)

Fill	Fill	Fill	Fill	Fill	Fill	Fill	Fill	Crit	Crit	Crit	Crit	Crit	Crit	Con	Con	Daily	Key
Pres	Pres	Pres	Non	Pres	Pres	Pres	Non	Pres	Pres	Non	Pres	Pres	Non	Pres	Pres	Out	Out
Daily	Daily	Daily	Daily	Key	Key	Key	Key	Daily	Daily	Daily	Key	Key	Key	Daily	Key		
Daily Training Examples																	
1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	+1	-1
1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	+1	-1
0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	+1	-1
Key Training Examples																	
0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	-1	+1
0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	-1	+1
0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1	-1	+1

Figure A5. Training and test stimuli for simulations of Experiment 3. Fill = filler; Crit = critical; Con = contradictory; Pres = presented; Non = nonpresented; Out = output.

to have different output levels, for example, if some prior beliefs are stronger than others or some categories are more entrenched. For example, if the PK nodes for sleeping pills and cooked breakfasts have higher output values than the PK nodes for children's and adults' breakfast cereals, the prediction would be faster learning for incongruent features than for critical features. If the PK nodes for sleeping pills and cooked breakfasts have lower output values, the prediction would be slower learning for incongruent features than for critical features. Regardless of these parameter values, the additional knowledge simulation predicts that both incongruent and critical features will be learned faster than filler features.

The right panel of Figure A4 shows predictions for nonpresented features. The model predicts an advantage for both incongruent features and critical features over filler features. Whether the incongruent features or the critical features will be learned faster will depend on the relative output values for PK nodes for sleeping pills and cooked breakfasts versus children's and adults' cereals. The key difference between this simulation and the simulation for incongruent weighting is that the simulation for additional knowledge predicts an advantage for nonpresented incongruent features over nonpresented filler features, whereas the incongruent weighting predicts no advantage.

Predictions of Model for Experiment 3

The simulation of Experiment 3 used the training stimuli shown in Figure A5, which correspond to the training stimuli used in the experiment itself. The network was structured as in Figure 2, except that there were no incongruent input nodes. There were eight filler input nodes, six critical input nodes, and two contradictory input nodes. There were fixed weights between critical features and PK nodes to represent prior knowledge, as in the simulation for Experiment 1. The contradictory input nodes were like additional critical input nodes except that the prior knowledge went in the opposite direction of other features for that category. For example, when the Daily category was associated with children's cereals, the six critical features for the Daily category would have a fixed connection weight of

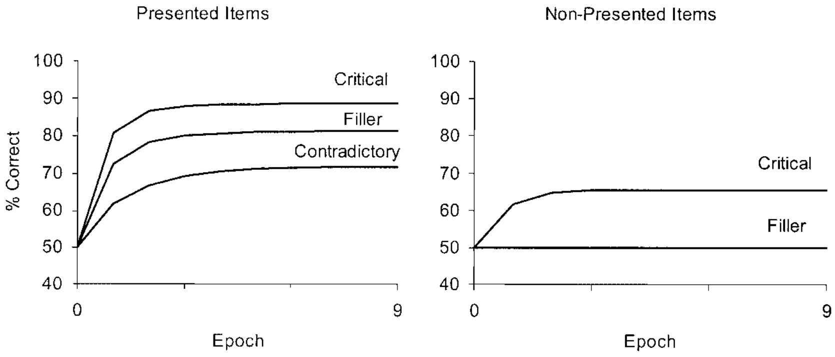

Figure A6. Predictions of the model for Experiment 3.

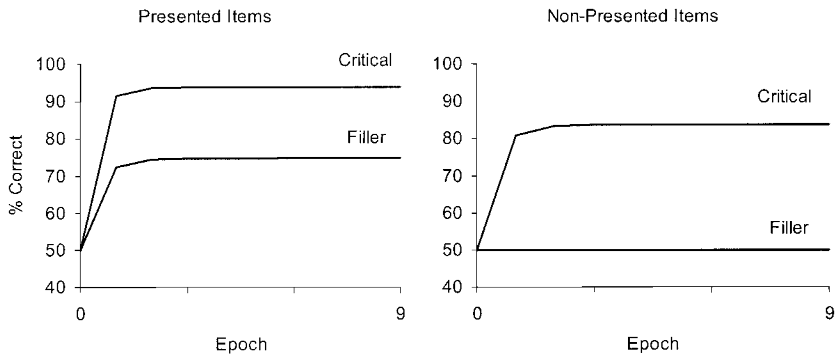

Figure A7. Predictions of the model for Experiment 3, with contradictory prior knowledge removed.

+1 with the children's cereal PK node and a weight of −1 with the adults' cereal PK node. The two contradictory features for the Daily category would have a fixed connection weight of −1 with the children's cereal PK node and a weight of +1 with the adults' cereal PK node.

The predictions are shown in Figure A6. On the left are predictions for presented features. The model predicts that critical features will have better performance than filler features, which in turn will have better performance than contradictory features. However, the prior knowledge effect, that is, the advantage of critical features over filler features, is attenuated. As an informal illustration of this attenuation, compare Figure A6 with Figure A2, which had a greater difference between critical and filler features. At the right of Figure A6 are predictions for nonpresented features. The model does predict some advantage for critical features over filler features, but again this advantage is attenuated, for example, compare critical features in Figure A6 versus in Figure A2.

This attenuation of prior knowledge effects due to contradictory features violating prior knowledge can be shown more directly by comparison with another simulation. This simulation was just like the main simulation for Experiment 3, except that the fixed connection weights from contradictory input features to PK nodes were all changed to 0. In other words, the contradictory features were presented in the same way but no longer contradicted prior knowledge. The predictions are displayed in Figure A7, which generally shows robust prior knowledge effects. That is, the difference between presented critical and filler features is attenuated in Figure A6 compared with Figure A7, and performance on nonpresented critical features is weak in Figure A6 compared with Figure A7. Therefore, it is the contradiction of prior knowledge that leads to the diminished prior knowledge effects.

Received September 18, 2003
Revision received February 3, 2004
Accepted February 10, 2004 ■