Journal of Experimental Psychology:
Learning, Memory, and Cognition
1998, Vol. 24, No. 3, 712-731

0278-7393/98/$3.00

Influences of Prior Knowledge on Selective Weighting of Category Members

Evan Heit
University of Warwick

Three experiments addressed how prior theories affect categorization, comparing the influence of theory-congruent versus theory-incongruent category members. Subjects observed descriptions of persons, some congruent with prior knowledge and some incongruent, then made transfer judgments. In Experiment 1, subjects were given a relatively long time to study each description, whereas in Experiment 2 study time was manipulated between subjects. In Experiment 3, learning was self-paced by each subject. It was found that, with enough study time, prior knowledge had 2 distinct influences. First, prior knowledge provided an initial representation, subsequently revised in light of new observations. Second, incongruent observations had more impact than congruent observations on categorization. In comparison, when study time was more limited, revision proceeded in a Bayesian manner, in that congruent and incongruent observations had equal impacts.

When a person learns about a new category, are all category members treated the same? Or do some category members have a greater impact on categorization than others? For example, imagine a tourist who is visiting a European country where, by stereotype, the inhabitants are expected to be very friendly. As the visitor learns about this category of people, do some observations have more influence than others? It might seem plausible that friendly people would be remembered better, because they fit in with expectations, and meetings with unfriendly people might be ignored or discounted. Alternately, surprising category members might have more of an impact, so that unfriendly people would have a disproportionate influence on categorization. A final possibility is that the visitor would learn in an unbiased fashion, in effect counting up the friendly and unfriendly observations without giving one kind of observation more weight than the other.

A good deal of experimental research has shown that categorization is indeed guided by expectations or prior knowledge (e.g., Hayes & Taplin, 1992, 1995; Heit, 1994, 1995; Murphy & Allopenna, 1994; Pazzani, 1991; Wattenmaker, 1995; Wisniewski, 1995; Wisniewski & Medin, 1994; see Heit, 1997, and Murphy, 1993, for reviews), but for the most part these studies did not address the issue of relative impact of different kinds of category members on categorization. That is, these past studies were not designed or analyzed with the intent of determining whether some category members, congruent with prior knowledge, have a different influence than other category members that are incongruent with prior knowledge. Yet considering the widespread extent of situations where categorization may be affected by some previous knowledge, the issue of which category members will have greater influence is critical for understanding how people make categorization judgments.

Indeed, it is possible to set out reasonable cases in favor of each of the three possibilities, namely (1) selective weighting in favor of theory-congruent category members, (2) selective weighting in favor of theory-incongruent category members, or (3) equal weighting of both kinds of observations. The next section presents arguments for these cases based on past results in related areas of psychology, and the following section addresses considerations of what might be optimal for categorization. At this point, the three possibilities are treated at a rather general level, with the goal of distinguishing among these broad accounts rather than specifying all their underlying processing mechanisms. However, the issue of underlying processes is taken up in more detail in the General Discussion.

Relevant Results in Categorization, Reasoning, and Memory

The Case for Favoring Congruent Category Members

Most of the past research on effects of prior theories on categorization has emphasized the facilitative effects of prior knowledge. In various ways, better performance seems to be associated with theory-congruent information rather than theory-incongruent information. For example, Murphy and Allopenna (1994) and Murphy and Wisniewski (1989) reported that when the features of a category were consistent with prior knowledge, performance was better (e.g., in terms of accuracy on test items) compared to situations where the features did not fit in with prior knowledge. Also, Pazzani

This research was presented at the 26th International Congress of Psychology, Montreal, Canada, August 1996, with travel support from the British Academy. This research was also supported by National Institute of Mental Health Grant MH10069 and by National Science Foundation Grant 91-10245.

I am grateful to Caren Jones, Koen Lamberts, Douglas Medin, Gregory Murphy, Robert Nosofsky, and Jeffrey Sherman 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. Electronic mail may be sent via JANET to E.Heit@warwick.ac.uk.

(1991) and Wattenmaker (1995) found that people learned a category's definition to a training criterion in fewer trials if the definition was congruent with prior knowledge than if it was incongruent with prior knowledge. Although these studies did not directly assess the influence of different category members on categorization, they do point to the idea of theory-incongruent information being poorly learned. It might be expected that poorly learned, incongruent category members would have a relatively low impact on categorization. Some research on reasoning and judgment also points to the general idea of people favoring cases to the extent that they fit with expectations or prior theories. Classic research on the confirmation bias (Wason, 1960; Mynatt, Doherty, & Tweney, 1977) has shown that in rule-discovery tasks, people tend to select test cases that would confirm a rule under consideration rather than choose test cases that might disconfirm the rule. Similarly, classic research on illusory correlation (Chapman & Chapman, 1967; Hamilton & Rose, 1980) has shown that people overestimate the association between events that are expected to co-occur. By analogy, when people make judgments about a new category they might select category members that fit with their prior theories and ignore category members that would disconfirm prior theories.

Some of the classic work on schematic effects on memory has pointed to the facilitative effects of prior knowledge. For example, Bransford and Johnson (1973) found that recall memory for a paragraph of text was improved by giving subjects a hint about relevant background knowledge (e.g., washing clothes). Similarly, Bower, Black, and Turner (1979) found that recall of events from stories was facilitated to the extent that the events were associated with an event script (e.g., going to a restaurant). Although other evidence from memory research, to be reviewed, does indicate that the picture is more complicated, at least some results do suggest that prior knowledge might facilitate memory for category members that are congruent with expectations. If category members that are incongruent with expectations do not obtain this facilitation, it might be thought that congruent category members would have a greater impact on categorization than incongruent category members.

The Case for Favoring Incongruent Category Members

Although it is possible to make a good case for favoring congruent category members, there would also be reasons to expect that incongruent category members might be favored instead. These reasons, however, would not directly come from past research on theory effects on categorization. Broadly speaking, past results in this area have been in the form of assimilation effects rather than contrast effects (see Heit, 1997, and Murphy, 1993, for reviews). That is, when subjects learned about new categories, what was learned seemed to be assimilated in the direction of their prior theories. Until now there have not been any reports of information that contrasts with prior theories being highlighted, facilitated, or otherwise having a greater influence on categorization.

Still, some research on category learning does bear on this issue indirectly. There is evidence that repeated presentations of category members have a diminishing marginal impact on categorization (Barsalou, Huttenlocher, & Lamberts, in press; Florian, 1992; Nosofsky, 1988). By analogy, a category member that fits in with prior theories is to some extent a repetition of what is already known. Thus, theory-congruent observations might have a diminished influence compared to theory-incongruent observations. Also, there is some evidence that category members are remembered better when they violate a rule of classification. Palmeri and Nosofsky (1995) found that recognition memory was better for category members that were exceptions to a categorization rule compared to category members that fit the rule. Again, these results serve as general evidence for the enhanced use of category members that do not fit in with expectations.

More generally in memory research, better performance on items that were unexpected is a fairly standard result. For example, Stangor and McMillan (1992) reviewed 54 experiments on how social stereotypes affect memory for descriptions of persons. The main question of interest was whether stereotype-congruent or stereotype-incongruent descriptions were remembered better. Although the experiments varied to some extent in their results, a meta-analysis showed a significant overall advantage for stereotype-incongruent stimuli, both for recall and recognition sensitivity. Graesser (1981) reviewed additional studies, particularly on memory for text, and the general result was higher recognition sensitivity for script-incongruent information. These results showing enhanced memory for incongruent observations are suggestive of a greater influence of incongruent category members in categorization, because better remembered category members could have a greater influence than poorly remembered category members.

The Case for Equal Weighting of Category Members

The most relevant evidence on whether theory-congruent or theory-incongruent category members have a greater influence on categorization is the one study that looked at this issue directly: Heit (1994). This work distinguished between two general ways that prior knowledge might affect category learning: selective weighting and integration. Selective weighting refers to favoring some observations over others, either favoring congruent over incongruent or vice-versa. In contrast, according to the integration account, prior theories serve as an initial anchor for category learning: The theory provides an initial category representation that is subsequently revised in light of observations of category members. In the example opening this article, the tourist exploring a European country would have an initial representation of the category of people in this country, even before setting foot in the country. This initial representation would be derived from knowledge about stereotypes and media reports and so on. The category representation would be revised as the tourist actually observes people in this new

country. That is, the initial representation based on prior knowledge would be integrated with the new observations. It would be possible for both integration and selective weighting processes to operate: Prior knowledge could provide an initial category representation, and subsequent observations could then be affected by selective weighting. Alternately, there could be integration and no selective weighting, or possibly selective weighting and no integration.

In brief, Heit (1994) provided evidence for integration at work, and equal weighting of congruent and incongruent observations. This evidence is reviewed in the next section. Whereas the results of Heit (1994) do provide a case of equal weighting, it was not clear how general these findings are. That is, with some variations in the learning situation, selective weighting of category members might be evident. One of the main purposes of the present article is to look at other conditions of learning, in an effort to find cases where selective weighting of observations might appear.

As already mentioned, there have been past results in memory research suggesting facilitation on congruent items as well as incongruent items. To some extent, whether congruent or incongruent items are favored could depend on the conditions of learning, leading to the variety of reported results. However, a different way of looking at this issue is that there may not be any selective weighting of observations. Heit (1993) conducted memory simulations in which memory traces for congruent and incongruent observations had equal weights. In addition, the simulations integrated memory traces for observations with memory traces representing prior knowledge. These simulations captured a number of trends in this area of memory research, providing an account for why memory will sometimes be better for congruent items and sometimes better for incongruent items. It is important to note that these simulations did not include any selective weighting of observations, and indeed Heit (1993) showed that selective weighting of observations made the simulations perform worse. Hence, there is some evidence that equal weighting of congruent and incongruent observations is a viable way to explain a range of results in memory tasks. Simply because test performance is better for a particular kind of stimulus, one should not assume that this kind of stimulus was selectively weighted.

Optimality Arguments

In addition to these arguments based on past results, it is possible to put forth cases on the basis of optimality. Philosophers and psychologists (Peirce, 1931–1935; Keil, 1989; Murphy & Medin, 1985) have argued that in many categorization tasks, particularly in real-world situations, there is an extremely large number of information sources and a combinatorial explosion of possible category representations, for example, combinations of features to be learned. Hence, it would seem beneficial to constrain the learning problem by favoring a small number of features or observations that seem most relevant. One way to enact this constraint would be to favor some category members over others, such as favoring theory-congruent category members over incongruent members. Doing so could simplify categorization by reducing the number of category members to be learned, and might possibly benefit categorization by filtering out irrelevant or noisy information. (This proposal is also consistent with the schema-as-filter hypothesis in memory research; e.g., Alba & Hasher, 1983.) Thus, selective weighting in favor of theory-congruent category members could be useful.

It is also possible to argue that it would be useful to weight incongruent category members more heavily than congruent category members. Perhaps in some situations, such as exploring new environments, it is beneficial to learn rapidly what is unusual in the new context. Theory-incongruent observations, because of their unfamiliarity, may require immediate attention. In contrast, theory-congruent observations may be less urgent because prior knowledge can be relied on for these cases. Also, theory-incongruent observations may simply seem more interesting or more likely to have implications for revising general knowledge. Therefore, it could be quite useful to favor theory-incongruent category members.

Finally, it is possible to argue that it would be optimal to weight congruent and incongruent observations the same. Category learning can be thought of as a statistical estimation problem, where the goal is to learn a parametric description of a category (such as that 85% of the people in some country are friendly). To learn an accurate representation of a category, it would be desirable for the learning process to be unbiased, and to converge in the asymptote. That is, as a person makes more observations, the person's category representation should become increasingly accurate, and in the long run the category representation should be as veridical as possible. It is notable that Bayesian estimation procedures (Box & Tiao, 1973; Raiffa & Schlaifer, 1961) have these characteristics, and these procedures are often considered normative. If congruent and incongruent category members are weighted equally, then the learner would have a chance of eventually forming an accurate category representation. In contrast, if the category learning process involves selective weighting, either in favor of congruent or incongruent category members, then what will be learned is an exaggeration or caricature of the category, with some category members grossly overrepresented.

Predictions for Varieties of Weighting

In an effort to distinguish among these three possibilities (equal weighting, congruent weighting, and incongruent weighting), Heit (1994) conducted a series of experiments that simulated—in schematic form—the experience of visiting a new city and observing people there. This situation is a quintessential example of prior knowledge influencing category learning and categorization, because there are completely novel categories of people to be learned but there is also extensive prior knowledge that is relevant, such as stereotypes of people in other places. First, subjects saw training examples consisting of featural descriptions of people in a novel city called City W, in an observational learning procedure. In effect, people were learning about

contextualized categories, such as shy people in City W and happy people in City W. Then subjects were asked to make transfer judgments about additional people from City W. For example, subjects were asked to judge the conditional probability that another person from City W, who avoids parties, would fall in the category of shy people.

The transfer judgments are best described in terms of two experimental variables. First, the proportion of times that a description had appeared in a category in City W, in the training phase, was examined at five levels from 0% to 100%. For example, the proportion of people who avoided parties in City W that had appeared in the *shy* category was 0%, 25%, 50%, 75%, or 100%. Second, half of the test questions involved a pairing that was congruent with prior theoretical knowledge, such as people who avoid parties being shy. Also, half of the test questions involved a pairing that was incongruent with prior knowledge, such as persons smiling more than average falling in the category of people who are unhappy.

Heit (1994) argued that equal weighting would be evidenced by *independent* influences of these two experimental variables on subjects' judgments, and selective weighting would be indicated by a particular pattern of *interactive* effects. To explain the predictions for equal weighting, it is useful to present a simple Bayesian model of probability judgment, which implements an integration process with equal weighting. (This model is shown to be related to other important models of categorization in later sections of this article.) Equation 1 is used to estimate a proportion, such as the proportion of times that an item with some description will be in a particular category. The variable q represents a prior estimate for this proportion, derived from previous knowledge. The variable G indicates the strength of belief in this prior estimate. The variable p indicates the proportion that has actually been observed after N cases. When N is small, that is, when few observations have been made, the estimated proportion, P_N, depends mainly on prior knowledge. But as N increases, the estimated proportion increasingly reflects the observed proportion, p (that is, the estimator is unbiased and it converges in the asymptote).

$$P_N = \frac{Np + Gq}{N + G} \tag{1}$$

It should be clear from Equation 1 that the prior estimate, q, and the observed proportion, p, have independent influences on judgments, because P_N is simply a linear combination of these two sources of information. Likewise, Heit (1994) argued that an integration process with equal weighting of observations would lead to independent effects of prior knowledge and observations. Figure 1A illustrates the estimated proportions predicted for this account. The initial prior estimate, q, is either 90%, representing a theory-congruent pairing, or 10%, representing a theory-incongruent pairing. The initial strength of the estimate, G, is set to 10 for the purpose of illustration. Along the x-axis of each graph, the observed proportion of category membership, p, is varied from 0% to 100%. The predictions are shown after $N = 10$ presentations. The parallel pattern of lines in Figure 1A illustrates the prediction that prior knowledge and observations have independent effects on judgments of likelihood.

Now, within this formal framework it is also possible to represent selective weighting of observations. Again, a learner could start with an initial estimate, q, having strength G. But rather than each observation having the same impact on the running average of p, some observations might have greater weights than others. For example, theory-congruent observations, such as people who avoid parties who are shy, might have more influence on later judgments than theory-incongruent observations, such as people who avoid parties but are not shy.

Figure 1B shows the predicted judgments when theory-congruent observations have 2.5 times the weight of theory-incongruent observations; for example, people who avoid parties and who are shy have 2.5 times the influence of people who avoid parties but are not shy. (Equation 1 itself would not be used to model selective weighting. See Appendix A for details of how the predictions of the model, with a selective weighting component, are derived.) In this figure, the influence of prior knowledge, as measured by the difference between the $q = 90\%$ line and $q = 10\%$ line, depends on the observed proportion, p. In particular, the difference is greater near the middle of the range ($p = 50\%$) than near the extremes ($p = 0\%$ and $p = 100\%$). Intuitively, selective weighting of observations should make a greater difference when the observations are mixed (50% congruent and 50% incongruent) compared to when the observations are unmixed (e.g., 100% congruent), because when the observations are, for example, all congruent, the relative influence of congruent to incongruent observations should have little impact. If hardly any of the observations are incongruent, it should not matter much whether incongruent observations are ignored. Hence, the prior knowledge effect, or difference between the $q = 90\%$ line and the $q = 10\%$ line, will depend on the observed proportion, p.

Finally, Figure 1C illustrates the predictions when theory-incongruent category members have a greater influence on categorization. Here, the model makes predictions for the case where theory-incongruent observations have 2.5 times the weight of theory-congruent observations. Again, there is a predicted interaction between the two variables. The predicted influence of prior knowledge is least when $p = 50\%$. Here, intuitively, the effect of selective weighting of incongruent observations runs in the opposite direction of initial beliefs, so the net effect of prior knowledge is especially reduced when the observations are mixed (50% congruent and 50% incongruent).

The results of Heit (1994) clearly favored the predictions for equal weighting (see also Heit, 1995). For example, Figure 2 shows the outcome of Experiment 2 of Heit (1994), with the average responses indicated by points on the graph and the predictions of an exemplar model of categorization indicated by lines. The lines labeled as congruent refer to conditional probability judgments between features that are congruent with each other according to prior knowledge, such as a judgment of the likelihood that someone who

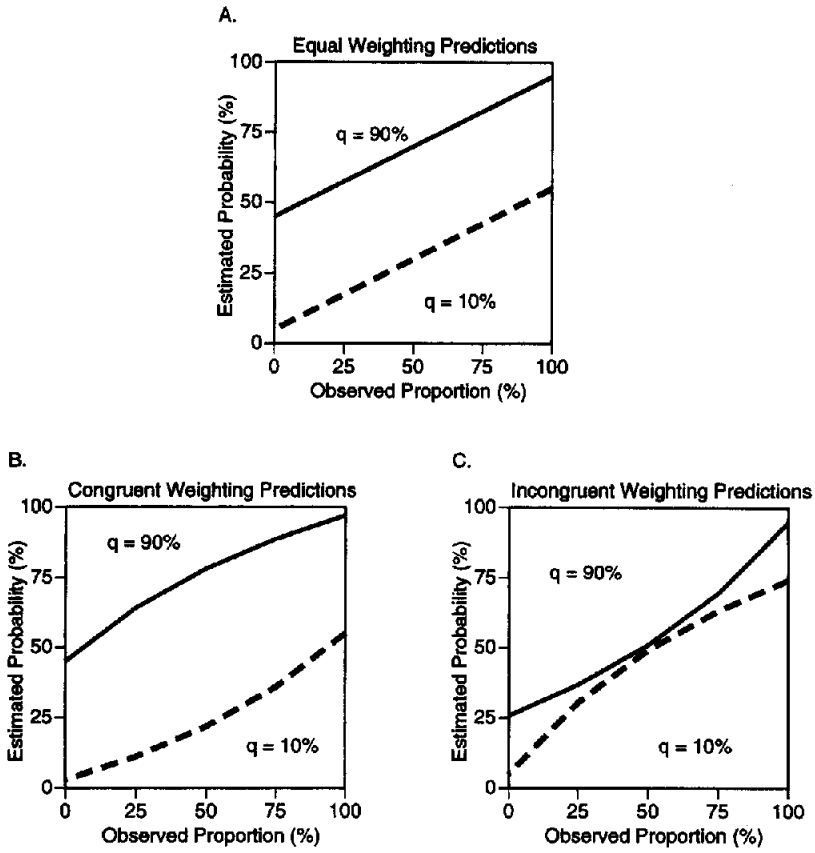

Figure 1. Sample predictions for equal weighting, for selective weighting of congruent category members, and for selective weighting of incongruent category members.

avoids parties will be in the *shy* category. Likewise, the lines labeled as incongruent refer to probability judgments between features that are incongruent with each other according to prior knowledge, such as a judgment of the likelihood that someone who smiles a lot is in the category of unhappy people. The *x*-axis indicates the observed proportion of category membership, in the training trials, for the test question. There was no evidence at all for selective weighting of observations, either for extra weighting of theory-congruent observations or for extra weighting of theory-incongruent observations.

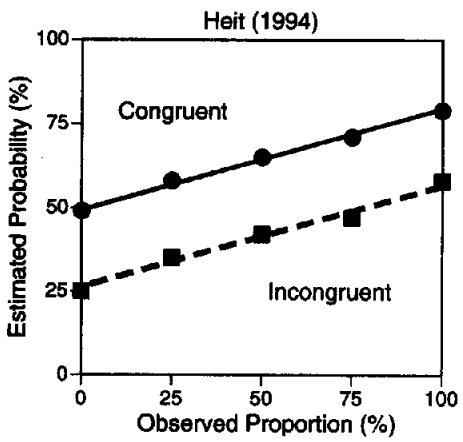

Figure 2. Results of Heit (1994, Experiment 2) and predictions of integration model with equal weighting.

Overview

The goal of these experiments was to examine the different ways that prior knowledge affects categorization, with particular emphasis on seeking situations where some category members might have greater influence than others. The present experiments were an attempt to vary the conditions of learning from the methods used in Heit (1994, 1995), which found equal weighting. Although there are good reasons to justify equal weighting, as previously mentioned there are also cases to be made for expecting selective weighting, in favor of either congruent observations or incongruent observations.

One possibility is that selective weighting processes are optional or otherwise contingent on the circumstances of learning. Selective weighting might require additional cognitive resources and thus depend on the nature of the stimuli, how the stimuli are presented, or the goals of the learner. The present experiments differed from the previous series on a

number of dimensions: encoding time, number of presented individuals, number of features per individual, and processing goal during learning. The general idea was to vary the experience of the subject in several salient ways, with the possibility that one or more of these experimental variables could have an influence on selective weighting. The dimensions to be investigated were motivated in part by past results on stereotype effects on person memory, using recall and recognition tasks. The meta-analysis conducted by Stangor and McMillan (1992) showed robust influences of these experimental variables. However, these past experiments had not focused directly on selective weighting of observations, and they had often used quite different experimental procedures. Thus, it was not possible to predict in advance how these experimental variables might specifically influence selective weighting in categorization.

Experiment 1

This experiment followed the general scheme of Heit (1994). The subjects observed descriptions of people in a fictional city called City W, and in effect they were learning about contextualized categories such as shy people in City W and college-educated people in City W. The transfer tests involved giving subjects partial descriptions of additional people from City W and asking for categorization judgments. More specifically, the task was to estimate the likelihood that a City W resident with a particular description, for example, attending parties often, would fall in a category of interest, for example, shy people.

It was suspected that encoding time would be a critical factor in how theories would influence categorization. In corresponding work on person memory (Bargh & Thein, 1985; Stern, Marrs, Millar, & Cole, 1984), it has been shown that prior knowledge effects are enhanced, and that processing is seemingly more elaborate, when learning occurs at a slow pace compared with when learning is speeded. Likewise, presenting fewer stimuli rather than more stimuli seems to enhance some stereotype effects (Rothbart, Fulero, Jensen, Howard, & Birrell, 1978; Stangor & Duan, 1991). Thus, it was hypothesized that slowing the pace of learning and reducing the number of stimuli to be observed might reduce processing load (Macrae, Hewstone, & Griffiths, 1993), and hence permit optional selective weighting processes to take place. Whereas the Heit (1994, 1995) experiments displayed descriptions of up to 200 persons, for 3.5 s each, Experiment 1 used only 40 stimulus descriptions, displayed for 16 s each. In terms of the range of memory experiments reviewed by Stangor and McMillan (1992), the Heit (1994, 1995) experiments were relatively fast in terms of presentation rate, whereas Experiment 1 was relatively slow-paced. Similarly, the earlier experiments would be at the high end in terms of number of individuals presented, whereas the present experiment represents a substantial move toward the lower end.

The stimuli in Experiment 1 were intended to be somewhat more complex than those in the previous series, again as an effort to encourage more elaborate processing. Whereas the Heit (1994, 1995) descriptions each consisted merely of two features, in this experiment the stimuli each included a person's name and four personal characteristics. The intent behind this change was to make the stimuli somewhat more interesting without overburdening memory resources. Again, using more interesting stimuli might encourage subjects to apply optional or effortful selective weighting processes. Both numbers of features per individual, two and five, would fall in the middle of the range in terms of Stangor and McMillan's (1992) review of memory experiments.

Finally, in Experiment 1 an attempt was made to influence the processing goals of the learner. Following research on person memory showing differences between memorization instructions and impression formation instructions (e.g., Hamilton, Katz, & Leirer, 1980; Srull et al., 1985), these two sets of instructions were compared. Again, it was hypothesized that impression formation instructions might encourage more elaborate processing and, thus, selective weighting effects. Memorization instructions might encourage simple, rote learning strategies, whereas impression formation instructions would seem to encourage a synthesized analysis or evaluation.

Method

Subjects. Forty-eight Northwestern University undergraduates participated; they received course credit or a small payment. All subjects in these experiments were recruited in the same manner. Half of the subjects in Experiment 1 received memorization instructions before learning, and half received impression-formation instructions. This experiment was run with a yoked design, to make the two conditions as similar as possible in terms of stimuli presented. Pairs of subjects, one in the memorization condition and one in the impression condition, saw exactly the same stimuli in the same order.

Stimuli. Each training example was a description of a person in terms of a name and four characteristics. For example, someone was named T. Kepler, was shy, did not attend parties often, bought expensive wine, and bought gourmet food. The 40 names (first initial and surname) were chosen at random from the Evanston, Illinois, telephone directory.

The training examples were derived from the descriptive terms shown in Appendix B. In the appendix, each couplet of four features comprises two pairs of opposites or complements. For example, *not shy* is the complement of *shy*, and *does not attend parties often* is the complement of *attends parties often*. The first and third item in each couplet were congruent with each other (e.g., *shy* and *does not attend parties often*), likewise the second and fourth item were congruent. The first and fourth items, as well as the second and third items, were incongruent (e.g., *shy* and *attends parties often*). The stimuli were pretested on other undergraduates to validate this manipulation of prior knowledge (see Heit, 1994, Appendix A).

The 40 training examples were organized in terms of five sets of eight examples. Two couplets of features were assigned randomly to each set. Each particular feature (e.g., *shy*) appeared four times, according to the following scheme. One set of examples contained information that was 100% congruent with prior knowledge. For example, there might be a total of four persons (with different

names) with the same description as T. Kepler, above. Likewise, there would be four persons who are not shy, do attend parties often, do not buy expensive wine, and do not buy gourmet food. Note that this description also contains congruent pairings of features. The remaining sets of examples contained 75%, 50%, 25%, or 0% theory-congruent information. For example, the 25% congruent set might appear as follows: one person with {generous, donates to charity, happy, smiles more than average}, three persons with {generous, does not donate to charity, happy, smiles less than average}, one person with {not generous, does not donate to charity, sad, smiles less than average}, and three persons with {not generous, donates to charity, sad, smiles more than average}.

What is most critical is the design of the test stimuli, with two within-subject variables. Each test question was a conditional probability judgment for a pair of features taken from a couplet in Appendix B, referring to the probability of one feature given another feature. The first experimental variable was whether the two features were congruent or incongruent with each other, according to prior knowledge. For example, a question referring to the conditional probability of a person being happy given that he or she smiles less than average would be incongruent according to prior knowledge. The second variable was the actual conditional probability, for the two features, of presentation during the study phase: 0%, 25%, 50%, 75%, or 100%. On test questions involving two congruent features, this percentage is equivalent to percentage of congruent presentations during the training phase, and on test questions involving two incongruent features, the percentage is equivalent to percentage of incongruent presentations during the training phase. Eight test questions were derived from each couplet; thus, there were 80 test questions.

Procedure. The procedure consisted of two parts, a training phase and a test phase. All information was displayed on a computer screen. Before the training phase, all subjects were told that they would see a number of descriptions of persons living in City W, a city located in Illinois. Subjects in the memorization condition were instructed to try to memorize these descriptions for a later test, whereas those in the impression-formation condition were instructed to form a general impression of City W, for a later test. The memorization instructions were nearly identical to those used by Heit (1994, 1995).

In the training phase, each participant saw 40 person descriptions displayed in a random order, one at a time for 16.0 s. There was a brief interval between displays (0.2 s), during which the computer screen was cleared. The person descriptions were presented as in the following example:

> T. Kepler has this description:
> shy
> does not attend parties often
> often buys expensive wine
> often buys gourmet food.

The order of these four features was determined randomly for each display, with the constraint that related features (e.g., *shy* and *does not attend parties often*) were kept adjacent to each other.

After the 20th training example had been displayed, participants were each given a 1-min rest period. Also, the training phase was followed by a 1-min break.

Next, in the test phase, subjects made 80 conditional probability estimates. The test questions for each subject were presented in a random order. These questions were worded as in the following example:

> Consider a person from City W with the following characteristic: shy
> How likely is it that this person would also have this characteristic? attends parties often.

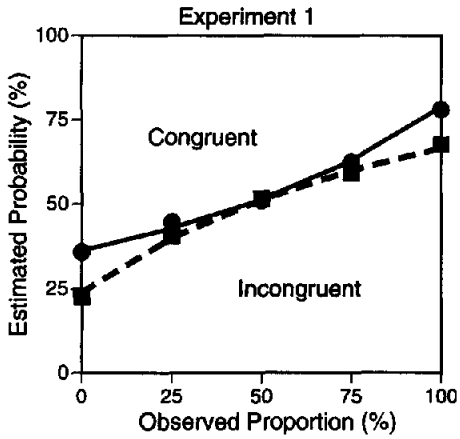

Figure 3. Results of Experiment 1 and predictions of mixed integration plus weighting model.

The subjects responded by typing integers on a scale from 0% to 100%. They were told to base their answers on what they inferred to be true of persons in City W after having seen descriptions of some of the residents of City W.[1]

Results

Observed probability estimates. The results of Experiment 1, in terms of the average response for each type of test question, are shown in Figure 3. (The results are averaged over the two instructional sets, which did not differ significantly.) Congruent test questions refer to conditional probability judgments between features that are congruent with each other according to prior knowledge, such as the conditional probability of a person who owns expensive running shoes being a jogger. Incongruent test questions refer to probability judgments between features that are incongruent with each other according to prior knowledge, such as the conditional probability of a person who smiles less than average being usually happy. This figure stands in contrast with the previous results of Heit (1994, 1995) that, as in Figure 1A, showed an independent contribution of prior knowledge for different levels of presentation or observed proportion. Figure 3 indicates a clear interaction between prior knowledge and what was presented in City W, showing the predicted pattern when incongruent category members are weighted more heavily than congruent category members. In particular, it appears that the prior knowledge effect, that is, the difference between congruent and incongruent test questions, is greatest for 0% and 100% congruent presentations, and that the prior knowledge effect

[1] It is useful to note that the training phase used an unsupervised learning procedure (as in Heit, 1992, 1994), in which subjects simply observed cases and learned without further feedback. In the test phase, each characteristic was treated (in the experimental design) as both a predicting feature and as a category label. For example, subjects were given the *shy* feature and asked to judge the likelihood of the *attends parties often* category, and were also given the *attends parties often* feature and asked to judge the likelihood of the *shy* category.

is diminished for mixed presentations (25%, 50%, and 75% congruent).

An analysis of variance (ANOVA) supported these observations. There was a main effect of congruent versus incongruent test question, $F(1, 46) = 4.81, p < .05, MSE = 1{,}821$, with higher probability judgments overall for congruent questions. Also, there was a main effect of presentation frequency, $F(4, 184) = 123.88, p < .001, MSE = 432$, with higher probability judgments when presented proportions were higher. Most critical, the interaction between these two variables was significant, $F(4, 184) = 8.12, p < .001, MSE = 185$.

As a further analysis of this interaction, difference scores were computed representing the prior knowledge effect, or difference between congruent and incongruent judgments, at a given level of presentation. In terms of Figure 3, these difference scores were equivalent to the signed distance between the two lines, taken at the 0%, 25%, 50%, 75%, and 100% levels of presentation. Trend analyses revealed that the difference scores showed a significant quadratic component as a function of level of presentation, $F(1, 46) = 14.94, p < .001, MSE = 290$. In other words, there was a statistically significant trend for the prior knowledge effect to be greatest at extreme levels of presentation (0% and 100%) and to diminish as presentations approach the middle of the range (50%).

Finally, the effect of instructional set (memorization or impression formation) and its interactions with the other two variables did not approach statistical significance (Fs < 1.0).

Model-based analyses. Next, the results were analyzed in terms of models of categorization. Unlike the ANOVAs, which looked at individuals' responses, these models were applied to average responses at the group level. The models (described more completely in Appendix A) were used to estimate the extent of integration and weighting processes within a Bayesian framework. The influence of integration processes was represented by q, the prior estimate of the proportion of times that a description would fall in a theory-congruent category, and G, the strength of this initial estimate. Critically, if there is no integration at work, then G will be estimated as zero. Thus, the model does not assume that integration must be taking place, but it does allow for that possibility. In addition, the relative weighting, W, of incongruent and congruent observations was estimated. The degree of weighting was measured as the ratio of the influences of incongruent versus congruent observations on judgments. If W is estimated to be 1 (or near 1), it will indicate that no selective weighting is taking place.

The categorization models also had another free parameter, s, indicating the degree of memory confusions (Medin & Schaffer, 1978). The value of s was in the range from 0 to 1, with higher values of s indicating worse memory. (See Appendix A for further details.) The use of the s parameter is something of a departure from the pure Bayesian framework, but previous results have already shown that people are somewhat suboptimal compared to a pure Bayesian account (e.g., W. Edwards, 1968; Heit, 1995). Allowing for some degree of imperfection of memory can explain results indicating slow probability revision or "conservatism." With this additional assumption, the Bayesian model is equivalent to the widely applied context model of classification (Medin & Schaffer, 1978). Most critical, introducing the s parameter does not alter the general qualitative predictions shown in Figure 1. Possible values of the free parameters were evaluated iteratively, to maximize the goodness of fit (in terms of a least-squares criterion) of the model to the average responses.[2,3]

Two models were evaluated for Experiment 1. A mixed integration and selective weighting model included both kinds of prior knowledge effects. In addition, a restricted model, including an integration process but no weighting process, was evaluated. In the restricted model, the W parameter was fixed at 1, so that congruent and incongruent observations had equal weights. When the mixed integration and weighting model performs significantly better than a pure integration model, there is evidence for a selective weighting process beyond any influence of an integration process.

First, the mixed integration and weighting model was applied to the data. Estimated parameter values and goodness-of-fit measures for the models are shown in Table 1. The parameter values indicate that the prior estimate, q, was .93, with this initial estimate having a strength, G, of 1.42. (The prior estimate for incongruent questions was simply $1 - q$, or .07.) Also, the estimated W parameter suggests that incongruent observations had 2.56 times the influence of congruent observations. These parameter values are consistent with the conclusions derived from visual inspection of Figure 3 and the ANOVAs above, namely, that there was some influence of integration and that subjects weighted incongruent observations more than congruent observations. The root mean square error (*RMSE*) of this model was quite low, .0090, indicating a very small average discrepancy between the model and the data. The predictions of the mixed integration and weighting model are shown as the lines in Figure 3, superimposed over the data points. On visual inspection, too, the correspondence between the data and the model predictions is quite close.

Next, a pure integration model was applied to the data. This model was restricted so that congruent and incongruent

[2] In practice, it is possible to reduce the number of free parameters in the model and still obtain quite good model fits for the present experiments. In particular, the q parameter could be set arbitrarily to 1, and by compensation, the s parameter allowed to rise a bit (as in Heit, 1994). However, it is useful to have separate estimates for q and s because the ideas of initial beliefs and memory confusions are conceptually distinct. Also, to fit a larger data set, as in Heit (1995), having separate estimates for q and s does lead to significant improvements in goodness of fit.

[3] Also, the average responses in each experiment were adjusted by a calibration parameter. Subjects in these experiments showed a slight lack of calibration, such that complementary probabilities added to somewhat more than 100% (see also Heit, 1992, 1994; Wallsten & Gonzalez-Vallejo, 1994). For example, the average response in Experiment 1 was 51.3%. To compensate for this lack of calibration, a value (here of .013) was subtracted from each response. These calibration values are listed in Tables 1, 2, and 3 for various applications of the models.

Table 1
Summary of Model Fits for Experiment 1

Model	Incon–con wtg. ratio (*W*)	Prior estimate (*q*)	Prior strength (*G*)	Memory confusion (*s*)	Calib.	*RMSE*
Int. + wtg.	2.56:1	.93	1.42	.26	.013	.0090
Integration	[1.00:1]	.83	0.87	.32	.013	.0272**

Note. Incon = incongruent; con = congruent; wtg. = weighting; Calib. = calibration; *RMSE* = root mean square error; Int. = integration. The value in brackets indicates a fixed parameter rather than an estimated value.
**$p < .001$, worse fit than mixed model.

observations would have equal influence. In effect, this model can only predict parallel lines when fitted to the data in Figure 3. The results for the pure integration model are summarized in Table 1. Most critical, the *RMSE*, .0272, is rather poor, three times the error of the mixed model. Indeed, the fit of the pure integration model is significantly worse than that of the mixed model, after taking into account that the pure model has one less free parameter than the mixed model, $\chi^2(1) = 11.09$, $p < .001$.[4] Thus, the weighting component of the mixed model is making a significant contribution.

It is also possible to fit a pure weighting model to the results, in which the *W* parameter is allowed to vary freely and the *G* parameter is fixed at zero (cf. Heit, 1994). For all the experiments in the present article, the pure weighting model performed strikingly worse than either of other models, in terms of statistical significance and the qualitative pattern of predictions. Thus, none of these results can be explained in terms of weighting processes unless some degree of integration is also allowed (see also Heit, 1994).

Finally, the results were examined for consistency across stimuli, that is, across the 10 couplets in Appendix B. In some ways it would be ideal to fit the models separately to the results for each set of stimuli, but each analysis would only be based on a small number of data points (about one-tenth of the total) and thus could have poor reliability. As an intermediate position between separate analyses for each couplet and the analyses above, which pooled over all 10 couplets, the couplets were split into three groups, reflecting weak, medium, and strong prior beliefs. The data on prior beliefs were obtained from another group of subjects (Heit, 1994, Appendix A), who were tested with a similar procedure to the present experiment, but without first participating in a study phase. These subjects simply made judgments about various pairings of features on the basis of their prior knowledge. For the three weakest couplets, the average difference between congruent test questions and incongruent test questions was 41%. For the three strongest couplets, the average difference between congruent test questions and incongruent test questions was 55%, and for the remaining four couplets, reflecting intermediate prior beliefs, the average difference was 49%. Using the data from the present experiment the mean results were obtained separately for these three groups of stimuli, breaking down the results in terms of congruent versus incongruent test questions and the five levels of presentation.

On visual examination, the patterns of results for all three stimulus groups were much like the curves in Figure 3, consistent with the prediction when incongruent category members have greater influence than congruent category members. Both the mixed integration-and-weighting model and the pure integration model were applied to the results for each stimulus group. Again, a better fit for the mixed integration and weighting model would serve as evidence for a distinct contribution of selective weighting. The mixed model fit significantly better than the pure model for the weak stimuli, $\chi^2(1) = 9.05$, $p < .01$, and for the medium stimuli, $\chi^2(1) = 8.82$, $p < .05$, but the improvement was not quite statistically significant for the strong stimuli, $\chi^2(1) = 2.32$. In sum, the pattern of results suggesting incongruent weighting was fairly consistent across groups of stimuli.

Discussion

The results of Experiment 1 suggest that people were learning about the contents of the new, contextualized categories in City W by (1) deriving an initial estimate based on prior knowledge, (2) revising this estimate in light of what was observed in the new category, and (3) being influenced more by incongruent observations than by congruent observations. The first two findings are consistent with previous results (Heit, 1994, 1995), but the third finding is quite novel from the perspective of some research on theory effects on category learning. Whereas past work (e.g., Keil, 1989; Keleman & Bloom, 1994; Murphy & Medin, 1985; Murphy & Wisniewski, 1989; Pazzani, 1991) has emphasized facilitation on theory-congruent information, here subjects appeared to be more influenced by new category members to the extent that they were inconsistent with previous theoretical knowledge.

These results were also dramatically different from the previous experiments on theory effects (Heit, 1994, 1995), which showed equal weighting. Although there were several

[4] The nested models were compared using the technique of Borowiak (1989). In brief, when model A is a nonlinear model with *a* free parameters estimated using a least-squares criterion, and B is a restricted version of this model with *b* free parameters, the likelihood ratio statistic is $\lambda = (RSS_A/RSS_B)^{(k/2)}$, where *RSS* is the residual sum of squares of the model and *k* is the number of data points to be predicted (here, 10). Borowiak shows that $-2 \ln (\lambda)$ has a chi-square distribution with $(a - b)$ degrees of freedom. (See Lamberts, 1994, for another application of this technique.)

methodological changes from the previous experiments, informal debriefings of the subjects provided hints as to what was critically different here. In particular, some subjects reported quite a bit of elaborate processing during the study phase as they tried to understand why some people in City W would be so odd. It seems that subjects tried to explain why the incongruent observations, such as shy people who attend a lot of parties, were in violation of their theories (cf. Chi, DeLeeuw, Chiu, & LaVancher, 1994; K. Edwards & Smith, 1996; Hastie, 1981; Srull & Wyer, 1989). As suggested in the introduction, incongruent observations may seem to be more interesting or important in terms of implications for revising knowledge.

Elaborate processing of the incongruent observations could plausibly lead to their enhanced influence at the time of categorization. Possibly, theory-congruent observations would not attract as much processing and thus would be less influential. It appeared that the relatively high amount of study time (16 s per instance) permitted this sort of self-explanation process, compared to previous experiments with less than 5 s of study time per item. In Experiment 2, the variable of study time per item was manipulated directly. The stimuli themselves (i.e., personal descriptions including a name and four other characteristics) were held constant from Experiment 1 to Experiment 2.

The variable of instructional format (memorization vs. impression formation instructions) did not have a significant effect in this experiment, despite other evidence for the impact of this instructional manipulation (Stangor & McMillan, 1992). The difference in instructions in Experiment 1 was perhaps not a strong enough manipulation, amounting to a change of only a few sentences of instructions amidst a great deal of information presented to subjects. Alternately, perhaps this manipulation of goals truly does not have an influence on selective weighting. In Experiments 2 and 3, the memorization instructions were used, maintaining consistency with previous research (Heit, 1994, 1995).

Finally, it is notable that the ANOVA results, the categorization modeling results, and the pattern in Figure 3 provide converging and mutually supportive evidence. The categorization modeling uses the subjects' average responses, and it shows that a categorization model with integration as well as selective weighting of incongruent observations fits the data better than a model with no selective weighting. The visual evidence from Figure 3 also gives clear support to the predictions for integration plus incongruent weighting, again at the aggregate level of response. Finally, the ANOVA was based on individual subjects' data, and likewise it showed a main effect of congruence, consistent with integration, and a statistically significant trend for a diminished effect of prior knowledge when observations are mixed, the hallmark prediction for incongruent weighting. ANOVAs and similar statistical procedures are used to estimate the consistency of a result across individual subjects, and, hence, to draw generalizations about a wider population. Thus, the conclusions for Experiment 1 are well-supported at both the aggregate and individual-subject levels of description.

Experiment 2

This experiment was focused on the effect of study time, which was one of the more striking procedural variations between Experiment 1 and earlier experiments (Heit, 1994, 1995). In Experiment 2, study time per training example was manipulated between subjects. The premise of this experiment was that the selective weighting in favor of incongruent observations, observed in Experiment 1, might be optional or otherwise contingent on having enough processing time. That is, consistent with past work suggesting that people will try to explain anomalous observations (Chi et al., 1994; K. Edwards & Smith, 1996; Hastie, 1981; Srull & Wyer, 1989), it was expected that such explanation processes would require some time to be carried out. Thus, it was predicted that under slower paced learning conditions the results would replicate those of Experiment 1 in terms of showing the added influence of incongruent observations. In contrast, it was expected that under faster paced learning conditions there would not be time for explanation processes and the results would indicate equal weighting, thus replicating previous findings (Heit, 1994, 1995).

Method

The critical change from Experiment 1 was that half (46) of the subjects viewed the training examples at a fairly rapid pace (5 s per description), whereas the other half saw training examples at a slower pace (14 s per description). The faster pace was comparable to that of the Heit (1994) experiments (3.5 s), which showed equal weighting, with some allowance made for the higher number of characteristics presented per individual. The slower pace was close to that of Experiment 1 (16 s), which showed incongruent weighting. This experiment was run with a yoked design, so that pairs of participants, 1 in the slow condition and 1 in the fast condition, saw exactly the same stimuli in the same order. All subjects were given memorization instructions, because the instructional manipulation had no apparent effect in Experiment 1. Also, the 1-min rest period in the middle of the training phase was eliminated, because it might have reduced the impact of the slow-versus-fast training manipulation. Otherwise, Experiment 2 was carried out in the same manner as Experiment 1.

Results

Observed probability estimates. The results are shown in Figure 4, separately for the slow and fast conditions. The results of the slow condition, in Figure 4A, resemble the results of Experiment 1 in terms of the interaction between prior knowledge and presentation frequency. The effect of prior knowledge is diminished for mixed presentation levels (25%, 50%, and 75%) compared with unmixed presentations (0% and 100%). In contrast, the results of the fast condition, in Figure 4B, do not show this pattern of interaction between the two variables. The results for the fast condition appear to replicate a number of previous findings (Heit, 1994, 1995) showing independent effects.

For clarity, the findings for the two conditions are first presented in separate ANOVAs. In the slow condition, there was a main effect of congruent versus incongruent test

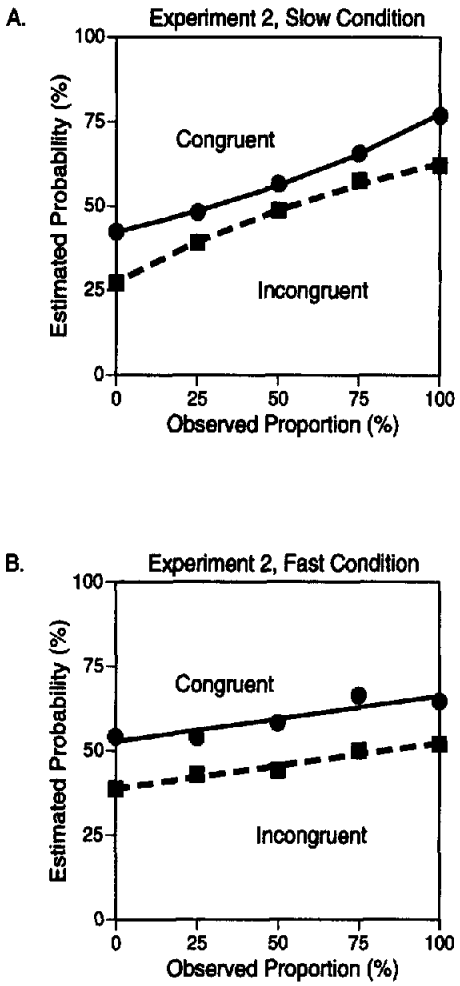

Figure 4. Results of Experiment 2, and predictions of mixed integration plus weighting model, for slow and fast conditions.

question, $F(1, 45) = 16.89$, $p < .001$, $MSE = 822$, with higher estimates for congruent questions. Also, there was a main effect of presentation frequency, $F(4, 180) = 59.76$, $p < .001$, $MSE = 295$, with higher estimates when observed proportions were higher. More critical, the interaction between the congruent–incongruent variable and the presentation frequency variable was significant, $F(4, 180) = 3.42$, $p < .01$, $MSE = 93$. As in Experiment 1, difference scores were computed at the five different levels of presentation, indicating the size of the prior knowledge effect. Again, these difference scores showed a significant quadratic trend as a function of presentation level, $F(1, 45) = 6.35$, $p < .05$, $MSE = 194$. In other words, as in Experiment 1, there was a statistically significant trend for the prior knowledge effect to be greatest at extreme levels of observation (0% and 100%) and to diminish as the presentation level approached the middle of the range (50%).

In the fast condition, there was a main effect of congruent versus incongruent test question, $F(1, 45) = 10.68$, $p < .01$, $MSE = 1{,}977$, with higher judgments for congruent questions. Also, there was a main effect of presentation frequency, $F(4, 180) = 26.91$, $p < .001$, $MSE = 99$, with higher judgments when presented proportions were higher. However, there was no evidence of a significant interaction between the congruent–incongruent variable and presentation level ($F < 1$). Unlike in Experiment 1 and in the slow condition of Experiment 2, there was no indication of the characteristic pattern of incongruent weighting, that is, smaller prior knowledge effects for mixed presentations than for unmixed presentations.

Next, the findings are presented from a three-way ANOVA using fast versus slow condition as a third, between-subjects variable, allowing all the results to be considered together. There was a main effect of congruent versus incongruent test question, $F(1, 90) = 24.73$, $p < .001$, $MSE = 1{,}400$, and a main effect of presentation frequency, $F(4, 360) = 85.85$, $p < .001$, $MSE = 197$. However, there was no main effect of fast versus slow condition ($F < 1$). Likewise, there was no significant interaction between congruence and fast versus slow ($F < 1$). There was an interaction between presentation frequency and fast versus slow, $F(4, 360) = 17.19$, $p < .001$, $MSE = 197$, suggesting that the presentation frequency variable had more influence in the slow condition than in the fast condition. Finally, there was no significant three-way interaction ($F < 1$). This last outcome, although inconclusive, was somewhat surprising, because it did not provide evidence that the two-way interaction between congruence and presentation frequency differed between the slow and fast conditions. Although in the separate ANOVAs, this two-way interaction was evident in the slow condition and not evident in the fast condition, still in the combined ANOVA there was no positive evidence that the interactions differed between the slow and fast conditions. This point is addressed further in the *Discussion* section, after relevant results from model-based analyses are presented.

Model-based analyses. Next, the categorization models were applied to the results from the slow and fast conditions. The fits of the models are described in detail in Table 2, and only the critical findings are noted here. For the slow condition, the mixed integration and weighting model fit the data extremely well, with an *RMSE* of only .0050. The predictions of the mixed model are shown superimposed over the data points in Figure 4A. As in Experiment 1, the estimated parameter values indicate a contribution of an integration process (a prior estimate of .94 with a strength equivalent to 1.95 examples) and selective weighting in favor of incongruent observations (with a ratio of 2.08:1). Also, as in Experiment 1, the mixed model fit the data significantly better than the pure integration model with equal weighting, $\chi^2(1) = 12.21$, $p < .001$. These model-based analyses suggest that the results of the slow condition are best explained in terms of both an integration process and twice the influence for incongruent observations compared with congruent observations.

When the mixed integration plus weighting model was applied to the fast condition, the best estimate for the *W* ratio was 1.04:1, or approximately equal weighting of incongruent and congruent information. The mixed model made nearly the same predictions as the pure integration model, which has the relative weighting of incongruent and congru-

Table 2
Summary of Model Fits for Experiment 2

Model	Incon–con wtg. ratio (W)	Prior estimate (q)	Prior strength (G)	Memory confusion (s)	Calib.	*RMSE*
			Slow condition			
Int. + wtg.	2.08:1	.94	1.95	.30	.024	.0050
Integration	[1.00:1]	.83	1.89	.32	.024	.0169**
			Fast condition			
Int. + wtg.	1.04:1	.88	5.39	.52	.025	.0167
Integration	[1.00:1]	.80	6.91	.46	.025	.0167

Note. Incon = incongruent; con = congruent; wtg. = weighting; Calib. = calibration; *RMSE* = root mean square error; Int. = integration. The values in brackets indicate fixed parameters rather than estimated values.
**$p < .001$, worse fit than mixed model.

ent observations fixed at a 1:1 ratio. Hence, the two models fit the data about equally as well, with hardly any difference in goodness of fit, $\chi^2(1) = 0.01$. The predictions of the mixed model are shown superimposed over data points in Figure 4B. Although the model fit is quite adequate, it is somewhat less impressive for the fast condition compared with the slow condition. The worse fit of the model here (*RMSE* = .0167) may reflect the difficulty that subjects had with learning in the fast condition, perhaps leading to unaccounted-for trends in the data due to guessing. Supporting this inference are the fairly high estimated values for the s and G parameters that suggest, respectively, poor overall memory and high reliance on prior knowledge rather than what was actually observed during the experiment.

Finally, the results were examined for consistency across stimuli. As in Experiment 1, the stimulus couplets were grouped in terms of weak, medium, and strong prior beliefs. In the slow condition, the pattern of results for both weak and strong stimuli was much like the pattern in Figure 4A, indicating selective weighting in favor of incongruent category members. For the weak and strong stimuli, the mixed integration and weighting model fit significantly better than the pure integration model, $\chi^2(1) = 5.00$, $p < .05$, for weak stimuli, and $\chi^2(1) = 6.61$, $p < .05$, for strong stimuli. These outcomes indicate a distinct contribution of selective weighting. However, the visual pattern for medium stimuli did not seem to suggest incongruent weighting, and indeed the mixed model did not fit any better than the pure integration model, $\chi^2(1) = 0.00$. In sum, in the slow condition there was good evidence for incongruent weighting in both the weak and strong stimuli. The lack of evidence for weighting in the medium stimuli could simply reflect an artifact of selecting a smaller subset of the whole pool of data. It is notable that in Experiment 1 there was significant evidence for weighting in medium stimuli.

The results were analyzed in the same manner for the fast condition. Most critical, visual examination of the pattern of results for the weak, medium, and strong stimuli did not suggest the presence of selective weighting. Moreover, in all three cases, the mixed integration and selective weighting model did not fit the results better than the pure integration model, $\chi^2(1) = 1.41$, $\chi^2(1) = 0.03$, and $\chi^2(1) = 0.05$, for the three stimulus groups, respectively. In sum, the lack of evidence for selective weighting was consistent across the three groups of stimuli.

Discussion

Both the qualitative results and the model-based analyses indicate that when given enough study time, subjects were influenced more by observations of theory-incongruent category members than by observations of theory-congruent category members. The results for the slow condition (14 s per observation) are quite similar to the results for Experiment 1 (16 s per observation) in terms of showing selective weighting of incongruent information. That is, the model-based analysis showed that an account with selective weighting of incongruent observations fit the overall data better than a model with no selective weighting. Also, the ANOVA for this condition showed a significant interaction between the prior knowledge and observation variables, consistent with the predictions for incongruent weighting. Finally, the visual pattern of results in Figure 4A nicely shows a diminished influence of prior knowledge nearer to the midpoint of the range of presentation frequencies.

In contrast, the findings for the fast condition (5 s per observation) appear to replicate the results of previous experiments (Heit, 1994, 1995) in which study time was less than 4 s per observation, in terms of suggesting equal weighting of category members. That is, the overall model-based analysis estimated approximately equal influences of incongruent and congruent observations (actually a 1.04:1 ratio), with the selective weighting component not significantly improving the fit of the model. Indeed, this lack of evidence for selective weighting was consistently observed for stimulus materials reflecting weak, medium, and strong prior beliefs. Also, the ANOVA for the fast condition showed main effects of prior knowledge and presentation frequency, without showing an interaction between these two variables. This result is consistent with an integration process alone without any influence of selective weighting. In addition, the visual pattern of the data points in Figure 4B, although having a somewhat jagged appearance, does not seem to show the hallmark pattern of results for incongruent weighting. That is, the effect of prior knowledge does not seem to

be diminished near the middle of the range of presentation levels (50%) compared to the extremes (0% and 100%).

However, these conclusions for the fast condition should be interpreted with some caution. Subjects in this condition, not surprisingly, did report that the rate of stimulus presentation seemed rather fast, and a few reported some guessing in the test phase as well. The model fits for this condition are worse than those for the slow condition or for Experiment 1. Finally, the combined ANOVA for the whole experiment did not show a significant three-way interaction between the slow versus fast variable and the main interaction of interest, between presentation frequency and congruence. All of these points suggest that the data for the fast condition were somewhat noisy, possibly reflecting other factors such as guessing.

Still, a few key points remain clear. There was quite a bit of positive evidence for selective weighting in the slow condition, and no evidence for selective weighting in the fast condition. Although the ANOVAs including the fast condition were not conclusive with respect to the three-way interaction, the overall responses for the fast condition suggest about equal influence of congruent and incongruent observations. Thus, the variable of study time seems to be important in determining how prior knowledge affects category learning. In several experiments now where category members were presented for 5 s or less (including this experiment, as well as Heit, 1994, 1995) the results all suggested equal weighting. Moreover, the finding that with brief study times, people do not favor incongruent information compared to situations with longer study times is consistent with some previous results on stereotype effects on memory (Bargh & Thein, 1985; Stangor & McMillan, 1992; Stern et al., 1984).

Putting together the results of Experiments 1 and 2 with previous results (Heit, 1994, 1995), it appears that selective weighting of incongruent category members is contingent on the conditions of learning. It appears that with a slower paced learning phase, additional processing takes place. This processing leads to a greater influence of incongruent category members on subsequent judgments, relative to congruent category members. As in Experiment 1, subjects in the slow condition reported in debriefings that they had tried to come up with explanations for the incongruent cases during the learning phase.

Experiment 3

Experiment 2, like Experiment 1, showed that when subjects had enough time during study, incongruent category members had a greater impact on transfer judgments than did congruent category members. Experiment 3 was expected to provide additional support for this finding. In this experiment, the pace of learning was determined by each participant rather than by the experimenter. The self-paced learning procedure is arguably more representative of real-world learning than an experimenter-paced procedure. It was expected that, overall, the results for self-paced learning would be similar to the results in Experiment 1 and the slow condition of Experiment 2 (cf. Bargh & Thein, 1985). Although it might seem clear in Experiments 1 and 2 that people were more influenced by category members that contrasted with prior knowledge effects, the long line of previous studies in this area showing general facilitative effects of prior knowledge suggests that there would be some value in an additional replication of this finding.

More critical, the self-pacing procedure permitted analyses of the link between study time and subsequent judgments. It was expected that incongruent category members would receive more study time than congruent category members (cf. K. Edwards & Smith, 1996; Stern et al., 1984), and that subjects who showed this difference to a greater extent would also show a greater impact of incongruent category members on transfer judgments. It was anticipated that overall, there would be a positive relationship between study time and influence at the time of testing, following classic Ebbinghausian principles of memory (e.g., the total time hypothesis, which states that more study time leads to greater accessibility at testing). Hence, if incongruent stimuli are studied longer, then they should have a greater influence in the test phase.[5] An analysis of the relation between study time and influence at testing was not possible for earlier experiments, because presentation time was the same for each item and looking time (or processing time) was not measured.

Examining the link between study time and categorization judgments would serve as a test of the model-based analysis itself. That is, the model with selective weighting of incongruent observations predicts a characteristic pattern of results, shown in Figure 1C. For Experiments 1 and 2, the model was used as a tool of data analysis to estimate, from the categorization results, how much selective weighting took place. In Experiment 3, selective weighting of observations, at study, was also measured indirectly in terms of study time. If the predicted pattern of results appears when subjects are selectively studying incongruent category members more than congruent category members, and this pattern does not appear otherwise, then the model itself would receive a degree of support.

Method

The critical change from Experiment 2 to Experiment 3 was that the 43 subjects viewed the training examples in a self-paced manner. At the beginning of the training phase, subjects simply were instructed to read the descriptions "at a comfortable pace." After each training example was presented for 1 s, the computer displayed a message to "press any key to continue." Thus, subjects were required to spend a minimum of 1 s study time per item, with no set maximum. The total study time for each item was recorded by the computer. Otherwise, Experiment 3 was carried out with the same method as Experiment 2.

Results

Observed study times and probability estimates. First, the study times were analyzed. Because the distribution of

[5] However, it is possible that incongruent stimuli are more difficult to process than congruent stimuli. This difficulty might reduce any effect of additional study time for incongruent stimuli.

study times for individual examples was positively skewed, these analyses were performed on log-transformed data. However, results are presented after a transformation back to the original measurement scale. The average study time for theory-congruent training examples was 8.61 s and the average study time per incongruent example was 9.31 s, a difference of 700 ms per example. A paired t test indicated that the difference was statistically significant, $t(42) = 2.99$, $p < .01$. Furthermore, the difference between congruent and incongruent study times was fairly consistent across subjects, with 32 of 43 subjects spending more time studying incongruent items than congruent items. For some of the subsequent analyses, the subjects were divided into two groups: the 32 who studied incongruent items longer and the 11 who studied congruent items longer. Among those who studied incongruent items longer, the average time for congruent items was 9.00 s and the average time for incongruent items was 10.51 s. Among those who studied theory-congruent items longer, the average time for congruent items was 7.57 s and the average time for incongruent items was 6.54 s. Thus, the subjects who spent more time on incongruent items than on congruent items also spent more time studying overall, compared with the other subjects.

Next, the responses in the test phase were analyzed, first looking at the responses from all of the subjects. The average responses, shown in Figure 5, resemble those in Figures 3 and 4A, with an interaction again suggesting a greater influence of theory-incongruent category members compared to congruent category members. There was a main effect of congruent versus incongruent test question, $F(1, 42) = 10.10$, $p < .01$, $MSE = 969$, and a main effect of presentation frequency, $F(4, 168) = 115.31$, $p < .001$, $MSE = 263$, with higher judgments when presented proportions were higher. Most critical, the interaction between the congruent–incongruent variable and presentation level was significant, $F(4, 168) = 2.72$, $p < .05$, $MSE = 113$. As in the previous experiments, difference scores were computed at the five different levels of presentation, indicating the size of the prior knowledge effect. Again, these difference scores showed a significant quadratic trend as a function of presentation level, $F(1, 42) = 4.25$, $p < .05$, $MSE = 263$.

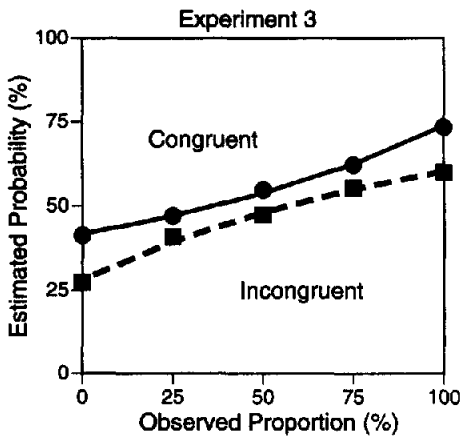

Figure 5. Results of Experiment 3 and predictions of mixed integration plus weighting model.

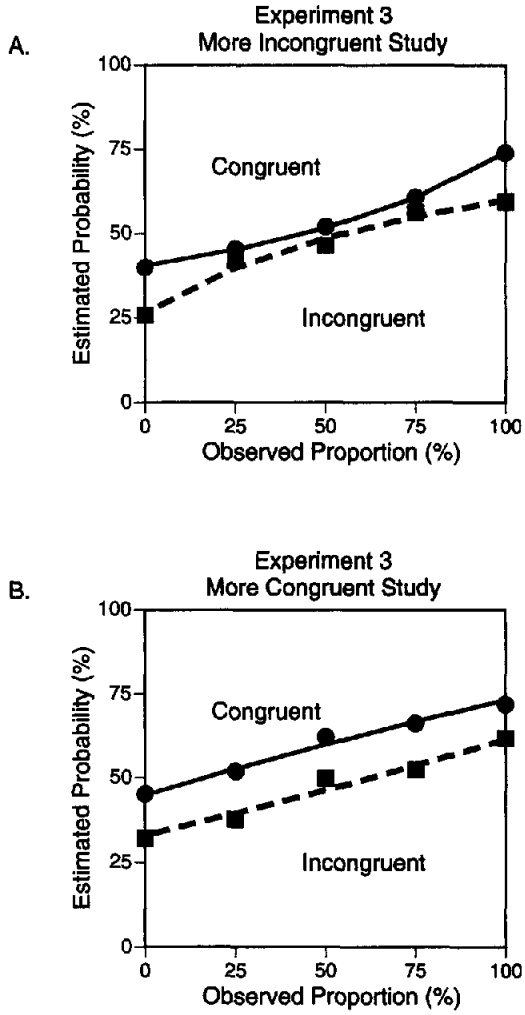

Figure 6. Results of Experiment 3 and predictions of mixed integration plus weighting model, shown separately for learners who studied incongruent category members longer and learners who studied congruent category members longer.

These analyses were repeated for the 32 subjects who actually spent more time studying incongruent training examples. Figure 6A shows their average responses; the results are quite similar to those shown in Figure 5, except that the attenuation of the prior knowledge effect near the midpoint (50%) is even more pronounced. There was a main effect of congruent versus incongruent test question, $F(1, 31) = 5.72$, $p < .05$, $MSE = 995$, and a main effect of presentation frequency, $F(4, 124) = 45.36$, $p < .001$, $MSE = 242$, with higher judgments when observed proportions were higher. The interaction between the congruent–incongruent variable and presentation level was significant, $F(4, 124) = 4.03$, $p < .01$, $MSE = 118$. Again, difference scores were computed at the five different levels of presentation, and these difference scores showed a significant

Table 3
Summary of Model Fits for Experiment 3

Model	Incon–con wtg. ratio (W)	Prior estimate (q)	Prior strength (G)	Memory confusion (s)	Calib.	*RMSE*
			Overall			
Int. + wtg.	2.57:1	.90	2.22	.32	.009	.0067
Integration	[1.00:1]	.90	1.49	.39	.009	.0178*
			More incongruent study time			
Int. + wtg.	4.32:1	.91	2.18	.29	.001	.0110
Integration	[1.00:1]	.80	1.69	.36	.001	.0261*
			More congruent study time			
Int. + wtg.	0.83:1	.88	2.16	.39	.031	.0162
Integration	[1.00:1]	.86	2.46	.37	.031	.0166

Note. Incon = incongruent; con = congruent; wtg = weighting; Calib. = calibration; *RMSE* = root mean square error; Int. = integration. The values in brackets indicate fixed parameters rather than estimated values.
*$p < .01$, worse fit than mixed model.

quadratic trend as a function of presentation level, $F(1, 31) = 6.36$, $p < .05$, $MSE = 264$, suggesting a greater impact of incongruent category members on judgments.

Finally, these analyses were performed for the 11 subjects who spent more time studying congruent training examples. (Of course, with only 11 subjects, these analyses do not have a great deal of power.) Figure 6B shows the average responses; these results suggest an independent effect of prior knowledge regardless of proportion of presentation. The main effect of congruent versus incongruent test question approached statistical significance, $F(1, 10) = 3.37$, $p = .09$, $MSE = 1{,}059$, and there was a main effect of presentation frequency, $F(4, 40) = 7.60$, $p < .001$, $MSE = 340$. There was no suggestion at all of a significant interaction between the congruent–incongruent variable and presentation level ($F < 1$).

Model-based analyses. Next, the categorization models were applied to the overall results, as well as separately to the two groups of subjects. The fits of the models are described in detail in Table 3, and only the main points are noted here. The mixed integration and weighting model fit the overall results quite well, with a *RMSE* of .0067. The predictions of the mixed model are shown as lines superimposed over the data points in Figure 5. As in Experiment 1 and the slow condition of Experiment 2, the estimated parameter values indicate a contribution of an integration process (initial estimate of .90 with a strength of 2.22) and selective weighting in favor of incongruent observations (with a ratio of 2.57:1). Also, the mixed model fit the data significantly better than the pure integration model, $\chi^2(1) = 9.72$, $p < .01$. These model-based analyses are suggestive of both integration and incongruent weighting processes at work. When the models were applied to the responses of the 32 subjects who spent more time studying incongruent observations, the conclusions were similar. (See the lines in Figure 6A for the predictions of the mixed model.) The most salient point about the model fits for this group is the very high degree of incongruent weighting, 4.32:1, suggesting that theory-incongruent observations had over four times the influence of theory-congruent observations.

For the 11 subjects who spent more time studying congruent items, applying the mixed integration and weighting model suggested greater use of congruent observations, that is, a 0.83:1 ratio comparing the impact of incongruent observations to congruent observations. (The predictions of the mixed model are shown superimposed over data points in Figure 6B.) However, the mixed model did not fit the data significantly better than the pure integration model, in which congruent and incongruent observations have the same weight, $\chi^2(1) = 0.25$. Thus, this evidence for congruent weighting is merely suggestive.

Finally, the results were examined for consistency across stimuli, using the data from all 43 subjects. As in the previous experiments, the stimulus couplets were grouped in terms of weak, medium, and strong prior beliefs. The pattern of results for both medium and strong stimuli was like the pattern in Figure 5, suggesting selective weighting in favor of incongruent category members. For the medium and strong stimuli, the mixed integration and weighting model fit better than the pure integration model, $\chi^2(1) = 5.20$, $p < .05$, for medium stimuli, and $\chi^2(1) = 4.75$, $p < .05$, for strong stimuli. However, the visual pattern for weak stimuli was not suggestive of incongruent weighting, and indeed the mixed model did not fit any better than the pure integration model, $\chi^2(1) = 0.00$. Thus, there was good evidence for incongruent weighting for both the medium and strong stimuli. There was no distinct evidence for selective weighting in the weak stimuli, but this lack of evidence was interpreted as a chance result reflecting a smaller pool of data, rather than indicating something systematically different about the weak stimuli. Indeed, Experiment 1 showed significant evidence for selective weighting for weak and medium stimuli, and Experiment 2 showed significant evidence for weighting in both weak and strong stimuli.

Discussion

The overall results of Experiment 3 replicate the findings of Experiments 1 and 2, showing a greater influence of

theory-incongruent category members than theory-congruent category members. Indeed, for the 32 subjects who spent more time studying incongruent observations, it was estimated that incongruent category members had more than four times the influence of congruent category members.

Also, the results establish a link between study time of category members and their subsequent influence on judgments. For the 32 subjects who spent more time studying incongruent category members than congruent category members, their average judgments indeed showed the pattern predicted for when incongruent observations have a greater influence. The remaining subjects did not spend more time studying incongruent category members, and for these subjects incongruent observations did not appear to have a greater influence on transfer judgments than congruent observations. Thus, this link supports the analyses in Appendix A for the effects of weighting incongruent category members. That is, when the Bayesian formula is modified to allow for incongruent category members to have a greater influence than congruent category members, it predicts results resembling those in Figure 1C. Study time is used as an indirect, converging measure of selective weighting. For the 32 subjects who actually spent more time studying incongruent category members, the results (in Figure 6A) do indeed resemble those in Figure 1C.

Finally, it is interesting to speculate about other differences between the 32 subjects who spent more time on incongruent items and the 11 subjects who did not. There was also an overall difference in study time between these two groups, with the group of 32 subjects spending about 3 s more time per item, averaged over congruent and incongruent items. Perhaps these subjects were more motivated than the others. Thus, the selective weighting of incongruent category members by this group may be attributed not only to extra study time for incongruent observations but also to more effortful processing in general for this group. This conjecture is consistent with the general idea that selective weighting processes are optional. Also, these subjects might have had stronger prior beliefs, and thus may have given the incongruent items more consideration because of their high perceived level of incongruence. The present data do not allow for discrimination between these two possibilities, that the slower subjects were more diligent and that the slower subjects perceived greater incongruence. Indeed, each explanation may have some merit. For example, some of the slower subjects may have been especially motivated or diligent, whereas others may have been more surprised by the incongruent stimuli, and for some subjects both explanations could be correct.

General Discussion

Although there are some good reasons to expect that theory-congruent category members might have a greater influence than theory-incongruent category members, the present experiments, as well as those of Heit (1994, 1995), have found no compelling evidence supporting this idea. Instead, it appears that under faster paced learning conditions there is equal weighting of category members, and that under slower paced learning conditions, there is an opportunity for optional or contingent processes to take place, leading to a greater influence of incongruent category members on categorization. The findings of Experiment 3, which used a self-paced learning task, also indicate a link between extra processing, particularly of incongruent category members, and an increased influence of incongruent category members at the time of judgment. It seems plausible, on the basis of subjects' reports as well as past research (Chi et al., 1994; K. Edwards & Smith, 1996; Hastie, 1981; Srull & Wyer, 1989), that the extra processing included some kind of explanation or conflict resolution for the incongruent descriptions.

However, the results taken together do represent something of a puzzle. When learning takes place in a fairly rushed manner (less than 5 s study time per observation), people appear to process information in a roughly Bayesian way. That is, congruent and incongruent observations have the same impact. The major advantage of equal weighting of congruent and incongruent information is that learning is unbiased in the asymptote, as illustrated by the Bayesian revision formula in Equation 1. As N increases, that is, as more observations are made, P_N approaches p, the actual observed proportion. In contrast, when people have ample study time (over 10 s per observation), then they act in a clearly non-Bayesian manner, by favoring incongruent observations over congruent observations. When learning is biased in this manner, P_N does not approach p in the asymptote. Instead, what is learned is something of a caricature of what is observed, with all of the unusual information being exaggerated. Surprisingly, subjects with more time to learn behaved less optimally than subjects with restricted study time! So is it better to learn in a rushed manner, so as to not be biased?

One way to resolve this puzzle is to consider further the task faced by the experimental participants. As mentioned in the introduction, it may be beneficial to give incongruent observations immediate consideration rather than safe and familiar congruent observations. In the short run (e.g., when first entering a new context), it could be quite reasonable to pay more attention to theory-incongruent category members. It might be said that in the short-term, learning about incongruent cases has a higher utility than learning about congruent cases. Furthermore, it seems reasonable that subjects in a psychology experiment would adopt a strategy that would be useful in short-term, real-world situations. It is conjectured that in longer term learning situations, beyond the span of typical psychology experiments, people might initially favor theory-incongruent observations, but this bias would diminish over time so that in the asymptote people could form accurate category representations.

Finally, although these experiments do provide a clear demonstration of incongruent weighting, it is useful to consider how general these results are. Over the present series of experiments (including Heit, 1994, 1995), several

variables have been manipulated, including study time, instructions to the subject, number of features per description, category size, and so on. It appears that study time is one critical variable in obtaining incongruent weighting effects. However, many other details of the experiments have been held constant, such as the cover story of seeing people in City W. Thus, the present results should be regarded mainly as a first clear existence proof for incongruent weighting, rather than as a wide-ranging conclusion for a variety of categories and tasks. It would be interesting in future research to examine these issues with other kinds of stimuli (e.g., artifact categories) and different experimental procedures.

Relation to Contrast Effects

All three experiments provided evidence for contrast effects, such that category members had more impact to the extent that they were incongruent with prior theories. Although contrast effects have been evident in other areas of psychology (e.g., Goldstone, 1995; Huttenlocher, Hedges, & Engebretson, 1996; James, 1890/1981; Manis, Biernat, & Nelson, 1991; Schwarz, Strack, & Mai, 1991; Wedell, 1995), this finding is unique within this area of categorization research.

The design of these experiments, in which observations and prior knowledge were manipulated factorially, was critical in showing the contrast effects. For example, consider Figure 5. If Experiment 3 had only used the 50% level of observed proportion, then the results seemingly would have indicated assimilation effects with no evidence for contrast effects, because these two data points merely show higher congruent judgments than incongruent judgments. It was only possible to find evidence for contrast effects by looking at the whole pattern of results (facilitated by analysis with formal models). Also, consider Figure 3. If Experiment 1 had only included a 50% observation condition, there would have apparently been no evidence for prior knowledge effects at all. Here, the assimilation effect, attributed to an integration process, is perfectly balanced by the contrast effect, attributed to selective weighting of incongruent information, so that the two data points fall exactly on top of each other, as if there were no effect of prior knowledge. More generally, the experiments illustrate the benefits of systematic, factorial design compared to designs with a smaller number of conditions.

It should be emphasized that the findings of contrast effects in these experiments are for *relative* contrast effects. There was an assimilation effect overall. That is, probability judgments on congruent test questions were higher overall than probability judgments on incongruent test questions, reflecting people's strong prior beliefs that certain pairs of features will go together. But in addition, the analyses revealed that contrasting observations had a disproportionate impact on updating these beliefs.

How Are Integration and Weighting Implemented?

The analyses in this article have provided evidence both for integration processes, in which an initial estimate is put together with new observations, and for weighting processes, particularly for selective weighting in favor of the incongruent observations. The mathematical models applied here are meant to be computational-level descriptions (Marr, 1982) of how people might combine prior knowledge and observations in category learning. They are, therefore, very general accounts that could be instantiated in a number of ways. For example, the prior knowledge about the category could be embodied as a prototype, a set of exemplars, a rule with some degree of associative strength, or as a pattern of activations within a connectionist network. Thus, the G parameter in Equation 1 could refer to some number of prior examples in memory (Heit, 1994) or simply to the degree of influence of prior knowledge, whatever its format. Likewise, the observations could be stored in memory by updating the prototype, storing exemplars, revising the strength of the rule, or training the network. To the extent that a prototype, exemplar, rule-based, or connectionist model of categorization embodies the Bayesian principle, in Equation 1, of independently combining prior knowledge and new observations, the model should predict a pattern of results like that in Figure 1A.

Although it is possible to construct models that violate this Bayesian principle, in practice many categorization models are consistent with this general idea. For example, Heit (1994) showed that an exemplar model, derived from Medin and Schaffer (1978), is equivalent to the Bayesian model with additional assumptions about memory confusions. (See also Nosofsky, 1990, 1991, for discussions of the relation between exemplar models and Bayesian models.) Heit (1993) briefly described a model that stores prior knowledge as a prototype rather than as exemplars but otherwise acts like the exemplar model (and the Bayesian model). Also, Heit (1995) described a simple associative-rule model and simple connectionist network model that embody the general idea of independent influences of prior knowledge and new observations. Finally, Anderson's (1991) rational model of categorization has at its core revision formulas such as Equation 1. Each of these models would make predictions resembling those in Figure 1A. Thus, the predictions for the integration process with equal weighting are quite general, considering the centrality of Bayesian-style updating to various models of categorization. Although there is a strong tradition in categorization research to address issues of representation (e.g., Smith & Medin, 1981), in the present article no conclusions are drawn about the representational format either of prior knowledge or new observations (see also Heit, 1997).

Just as the integration of prior knowledge with new observations could be implemented in several ways, selective weighting processes could likewise be implemented in a number of ways. For example, favoring some category members and ignoring others could be a willful, strategic process on the part of subjects. Alternately, selective weighting could be an unintentional side effect, such as a result of poor memory, or weak memory traces, for some observations. For example, if congruent category members are very poorly remembered, then they may have less influence on categorization judgments than other, incongruent, category

members that are well-learned. Broadly speaking, selective weighting processes might take place at encoding and retrieval. For example, if some category members are strategically ignored or discounted, this could take place at encoding or retrieval. The present experiments, showing a lack of selective weighting when encoding time is brief, suggest that the locus of at least some of the selective weighting in these experiments was at the time of encoding. In sum, the goal of this article has not been to specify a list of processing details but rather to address general principles (see also Heit & Barsalou, 1996) that have implications for a variety of categorization models.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203–231.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Bargh, J. A., & Thein, D. (1985). Individual construct accessibility, person memory, and the recall–judgment link: The case of information overload. *Journal of Personality and Social Psychology, 49,* 1129–1146.

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (in press). Processing individuals in categorization. *Cognitive Psychology.*

Borowiak, D. S. (1989). *Model discrimination for nonlinear regression models.* New York: Marcel Dekker.

Bower, G. H., Black, J. B., & Turner, T. F. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177–200.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis.* London: Addison-Wesley.

Bransford, J. D., & Johnson, M. K. (1973). Consideration of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing* (pp. 383–438). New York: Academic Press.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 73,* 193–204.

Chi, M. T. H., DeLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71,* 5–24.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.

Elliott, S. W., & Anderson, J. R. (1995). The effect of memory decay on predictions from changing categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 815–836.

Florian, J. E. (1992). Frequency effects on categorization and recognition. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 826–831). Hillsdale, NJ: Erlbaum.

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science, 6,* 298–304.

Graesser, A. C. (1981). *Prose comprehension beyond the word.* New York: Springer-Verlag.

Hamilton, D. L., Katz, L., & Leirer, V. (1980). Cognitive representation of personality impressions: Organizational processes in first impression formation. *Journal of Personality and Social Psychology, 39,* 1050–1063.

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology, 39,* 832–845.

Hastie, R. (1981). Schematic principles in human memory. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium.* Hillsdale, NJ: Erlbaum.

Hayes, B. K., & Taplin, J. E. (1992). Developmental changes in categorization processes: Knowledge and similarity-based models of categorization. *Journal of Experimental Child Psychology, 54,* 188–212.

Hayes, B. K., & Taplin, J. E. (1995). Similarity-based and knowledge-based process in category learning. *European Journal of Cognitive Psychology, 7,* 383–410.

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology, 24,* 341–380.

Heit, E. (1993). Modeling the effects of expectations on recognition memory. *Psychological Science, 4,* 244–252.

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. (1995). Belief revision in models of category learning. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 176–181). Hillsdale, NJ: Erlbaum.

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

Heit, E., & Barsalou, L. W. (1996). The instantiation principle in natural categories. *Memory, 4,* 413–451.

Huttenlocher, J., Hedges, L., & Engebretson, P. H. (1996, November). *Category effects on judgement: Assimilation and contrast.* Paper presented at the 37th Annual Meeting of the Psychonomic Society, Chicago.

James, W. (1981). *The principles of psychology.* Cambridge, MA: Harvard University Press. (Original work published 1890).

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Keleman, D., & Bloom, P. (1994). Domain-specific knowledge in simple categorization tasks. *Psychonomic Bulletin & Review, 1,* 390–395.

Lamberts, K. (1994). Towards a similarity-based account of compatibility effects. *Psychological Research, 56,* 136–143.

Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology, 23,* 77–87.

Manis, M., Biernat, M., & Nelson, T. F. (1991). Comparison and expectancy processes in human judgment. *Journal of Personality and Social Psychology, 61,* 203–211.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

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

Murphy, G. L. (1993). Theories and concept formation. In I. V. Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 173–200). London: Academic Press.

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

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

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science* (pp. 23–45). Chichester, England: Ellis Horwood.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment. *Quarterly Journal of Experimental Psychology, 29,* 85–95.

Nosofsky, R. M. (1988). Similarity, frequency, and category

representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.
Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, 34,* 393–418.
Nosofsky, R. M. (1991). Relation between the rational model and the context model of categorization. *Psychological Science, 2,* 416–421.
Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 548–568.
Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 416–432.
Peirce, C. S. (1931–1935). *Collected papers of Charles Sanders Peirce.* Cambridge, MA: Harvard University Press.
Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory.* Boston: Harvard University, Graduate School of Business Administration.
Rothbart, M., Fulero, S., Jensen, C., Howard, J., & Birrell, P. (1978). From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology, 24,* 237–255.
Schwarz, N., Strack, F., & Mai, H.-P. (1991). Assimilation and contrast effects in part–whole question sequences: A conversational logical analysis. *Public Opinion Quarterly, 55,* 3–23.
Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.
Srull, T. K., Lichtenstein, M., & Rothbart, M. (1985). Associative storage and retrieval processes in person memory. *Journal of Experimental Psychology, 11,* 316–345.
Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review, 96,* 58–83.
Stangor, C., & Duan, C. (1991). Effects of multiple task demands upon memory for information about social groups. *Journal of Experimental Social Psychology, 27,* 357–378.
Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111,* 42–61.
Stern, L. D., Marrs, S., Millar, M. G., & Cole, E. (1984). Processing time and the recall of inconsistent and consistent behaviors of individuals and groups. *Journal of Personality and Social Psychology, 47,* 253–262.
Wallsten, T. S., & Gonzalez-Vallejo, C. (1994). Statement verification: A stochastic model of judgment and response. *Psychological Review, 101,* 490–504.
Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140.
Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology, 28,* 274–328.
Wedell, D. H. (1995). Contrast effects in paired comparisons: Evidence for both stimulus-based and response-based processes. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1158–1173.
Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 449–468.
Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–282.

Appendix A

Description of Categorization Models

The basic formula for Bayesian estimation of a proportion is shown in Equation A1. Here, q represents the prior estimate of the proportion, G is the strength of this estimate, p is the observed proportion, and N is the number of observations. This formula can be applied to categorization, for example, by estimating the proportion of times that a particular description falls in a category of interest:

$$P_N = \frac{Np + Gq}{N + G}. \tag{A1}$$

Equation A1 represents the case of equal weighting; that is, each observation has the same impact on the estimate, P_n. Within this framework it is also possible to represent selective weighting of observations, as illustrated by Equation A2. Here, each observation, o_i, would have a value of 1 or 0, denoting whether or not the item falls into the category. For example, people avoiding parties who are shy would be assigned the value of 1, and people avoiding parties who are not shy would be assigned the value 0. Each observation is also assigned a weight, w_i, which indicates its impact on the running average. In the simplest case, all of these weights are 1, and then Equation A2 reduces to Equation A1. To implement congruent weighting, theory-congruent observations would be assigned higher weights than theory-incongruent observations. Likewise, to implement incongruent weighting, incongruent observations would get higher weights:

$$P_N = \frac{\sum_{i=1}^{N} w_i o_i + Gq}{\sum_{i=1}^{N} w_i + G}. \tag{A2}$$

Equation A3 shows a simplified version of Equation A2, for the situation where theory-congruent observations are assigned a weight of 1, theory-incongruent observations are assigned a weight of W, and q is greater than .5. This equation was applied to derive predictions for the integration-plus-incongruent weighting model on congruent questions, for example, the top line in Figure 1C. Note that there are Np congruent observations with a weight of 1, and $N(1\text{-}p)$ incongruent observations with a weight of W. It is possible to state similar equations to derive the remainder of the model predictions. Note again that when $W = 1$, that is when

congruent observations and incongruent observations have the same weight, Equation A3 reduces to Equation A1:

$$P_N = \frac{Np + Gq}{Np + WN(1 - p) + G}. \quad \text{(A3)}$$

When the models were applied to experimental data, allowances were made for some degree of imperfections in memory. It is well-established that people's probability revision tends to be conservative relative to Bayesian models (W. Edwards, 1968), suggesting some degree of suboptimality of performance. There are a number of possible ways to add memory confusions or forgetting to Bayesian models (see Anderson, 1991; Elliott & Anderson, 1995; Heit, 1995, Nosofsky, 1991). The particular method used was chosen for its compatibility with existing work on the modeling of categorization. That is, Equation A4 is equivalent to the widely applied context model of categorization, an exemplar-based account (Medin & Schaffer, 1978; see Heit, 1994):

$$P_N = \frac{Np + sWN(1 - p) + Gq + sG(1 - q)}{[Np + WN(1 - p) + G][1 + s]}. \quad \text{(A4)}$$

In Equation A4, it is assumed that dissimilar descriptions will be confused with each other to some extent, s, where $0 \leq s \leq 1$. Say that Equation A4 is applied to judgments about whether people who avoid parties will be in the *shy* category. The subject has made Np observations of people who avoid parties that fall in this category. For the present experiments, subjects also observed members of this category with different descriptions, such as $N(1\text{-}p)$ people who do attend many parties. These mismatching descriptions also have some impact, s, on judgments about people who do avoid parties. Also, these descriptions, of people who attend many parties but are shy, are theory-incongruent, so they have a weight of W. Thus, the ultimate impact of these theory-incongruent, mismatching descriptions is $sWN(1\text{-}p)$. Likewise, although there is prior knowledge, Gq, supporting the probability judgment, there is also some evidence against, $G(1\text{-}q)$, which is retrieved when $s > 0$. Note that when $s = 0$, that is, when memory is perfect, Equation A4 reduces to Equation A3. The influence of higher values of the s parameter is merely to reduce the overall sensitivity to what is in memory. Most critical, including the s parameter does not alter the qualitative predictions illustrated in Figure 1.

A final point of comparison between the Bayesian model and the context model of categorization is that the context model typically has been applied to choice proportions rather than to subjects' direct estimates of probability. It is assumed that this Bayesian model would also be useful in predicting choice proportions, although forced-choice procedures were not used in the present experiments. However, previous studies (Heit, 1992, 1994) applied variants of the context model to both choice-proportion and probability-estimate data and obtained quite good fits for both kinds of data.

Appendix B

Feature Couplets Presented in Experiments 1–3

Shy
Not shy
Does not attend parties often
Attends parties often

Jogs regularly
Does not jog regularly
Owns expensive running shoes
Does not own expensive running shoes

Travels two or more times per year
Travels less than two times per year
Has frequent flyer number
Does not have frequent flyer number

Has a college degree
Does not have a college degree
Higher income than average
Lower income than average

Often buys expensive wine
Does not buy expensive wine often
Often buys gourmet food
Does not buy gourmet food often

Stubborn
Not stubborn
Frequently gets into arguments
Does not get into arguments frequently

Mechanically inclined
Not mechanically inclined
Fixes things as a hobby
Does not fix things as a hobby

Generous
Not generous
Donates to charity
Does not donate to charity

Usually happy
Usually sad
Smiles more than average
Smiles less than average

Watches more TV than average
Watches less TV than average
Reads books less than average
Reads books more than average

Received October 16, 1996
Revision received August 18, 1997
Accepted August 26, 1997 ■