Memory & Cognition
2001, 29 (6), 828-837

Putting together prior knowledge, verbal arguments, and observations in category learning

EVAN HEIT
University of Warwick, Coventry, England

Two experiments addressed the novel issue of how people incorporate verbal arguments into category learning. In Experiment 1, at the start of learning, subjects were given verbal arguments, which had an influence equivalent to a fixed number of category members. In Experiment 2, subjects learned under slower paced conditions, and it was found that both prior knowledge and arguments had multiple effects on categorization: a fixed initial influence plus selective weighting of new observations. The results supported the idea that verbally presented arguments can be treated in a similar manner as other forms of prior knowledge, from the perspective of applying models of categorization.

Recent results have indicated that categorization is influenced not only by the observed members of a category but also by prior knowledge (e.g., Hayes, Taplin, & Munro, 1996; Kaplan & Murphy, 2000; Palmeri & Blalock, 2000; Spalding & Murphy, 1996; Wattenmaker, 1995; Wisniewski & Medin, 1994; see Heit, 1997, for a review). These studies have extended the range of categorization research beyond that of numerous experiments that have focused solely on the impact of observed category members. This paper is intended to extend the range of categorization research even further by looking at how verbal arguments are incorporated into category learning. For example, a job applicant being recruited by an ambitious company might make observations of the workplace and the people there, but these observations could be accompanied by verbal arguments about why it is a good place to work and why it is better than other places. These arguments could present information (such as regarding employment benefits at this company) that would not be available from direct observations. In some situations, arguments could be used to contradict people's prior knowledge. For example, trainee counselors could be taught that, contrary to popular belief, shy people often attend parties as an attempt to mask their shyness.

The specific approach in this paper will be to apply a mathematical model of categorization that has origins in the artificial category learning literature (Medin & Schaffer, 1978) and has been applied to some influences of prior knowledge on categorization (Heit, 1994, 1998, 2001). Can this existing model also address how verbal arguments affect category learning? The working hypothesis will be that arguments are processed the same way as other kinds of prior knowledge. That is, the information available at the start of learning will be subject to the same processing whether this information comes from people's long-term beliefs or whether this information was recently presented in paragraph form. One consequence of this hypothesis is that verbally presented arguments should have qualitative effects comparable to other forms of prior knowledge.

What cognitive processes are involved in putting together prior knowledge and observations of category members? From the work of Heit (1994, 1998), a theoretical account of some of these processes has emerged. This account distinguishes between integration and selective weighting processes. For illustration, imagine someone taking a new job at an extremely prestigious research institute. This person knows the reputation of the research institute and its members and now has an opportunity to make direct observations and form an accurate representation. The question is, How will prior knowledge (the reputation) and observations be put together? The integration process refers to how prior knowledge serves as an initial representation, which is subsequently revised as new observations are made. In the present example, the initial representation would be very favorable (e.g., it would be initially assumed that everyone is very hard-working). In Heit's (1994) model of integration, prior knowledge was represented as being equivalent to a fixed number of observations. In contrast, the weighting process refers to how new observations themselves are treated. Do some observations count more than others? For example, if people are expected to be hard-working, would an observation of a lazy person have more or less impact than an observation of a hard-working person? This is a crucial issue for virtually all models of categorization that work from a representation that stores or summarizes observed category members.

This research was supported by the Economic and Social Research Council and the Biotechnology and Biological Sciences Research Council of the United Kingdom. The author thanks Lewis Bott, Shellie Cross, Ros Shadlock, and Vivienne Turley for assistance in developing the stimuli and conducting the experiments. Correspondence should be addressed to E. Heit, Department of Psychology, University of Warwick, Coventry CV4 7AL, England (e-mail: e.heit@warwick.ac.uk).

It appears that integration processes are mandatory, in that they affect category learning even under difficult (e.g., fast) learning conditions. Weighting processes are optional and strategic. Under easier (e.g., slower) learning conditions, people can apply processes that lead to selective weighting in favor of category members that are incongruent with previous expectations. Thus, the lazy person at the research institute would have a greater effect on the emerging category representation than would a hard-working person. This account is closely linked to research on how stereotypes affect memory (e.g., Bargh & Thein, 1985; Hastie, 1980; Macrae, Bodenhausen, Schloerscheidt, & Milne, 1999; Macrae, Hewstone, & Griffiths, 1993). These studies suggest that an automatic processing mode is sufficient for ordinary observations, but more controlled, effortful processing is required for responding to unusual or incongruent observations.

The main purpose of this paper is to provide converging evidence for this theoretical account by applying it to the novel issue of how people incorporate verbal arguments into category learning. Information that is supplied before any observations are made should act as another kind of prior knowledge, and, hence, the account developed by Heit (1994, 1998) should apply to this information as well. Therefore, verbally presented arguments, like other forms of prior knowledge, should have different effects under faster learning conditions and slower learning conditions.

In both category learning experiments presented here, the subjects were presented with verbal arguments related to some of the categories, such as a statement explaining why some shy people might actually attend a lot of parties. In Experiment 1, subjects learned under relatively fast-paced conditions (3.5 sec per description). It was predicted that, relative to observations, verbal arguments would have a statistically independent effect on categorization, just as prior knowledge did in previous experiments (Heit, 1994), but working in the opposite direction. Experiment 2 used a much slower-paced learning procedure (16 sec per observation). Under similar slow-paced conditions, Heit (1998) found that prior knowledge has two distinct influences on categorization. First, prior knowledge provides an initial set of expectations, as represented in the integration model by a set of prior examples. Second, category members that are incongruent with prior knowledge are highlighted and have a greater influence on categorization than do theory-congruent category members. The hypothesis behind Experiment 2 was that verbal arguments would again act like prior knowledge, having two distinct effects on categorization.

EXPERIMENT 1

The purpose of Experiment 1 was to counteract the pattern of prior knowledge effects with verbal arguments that went against usual expectations. For example, if subjects read an explanation for why shy people might actually attend a lot of parties, then thinking about this explanation might serve to neutralize the effect of subjects' prior beliefs about shy people. In Experiment 1, the additional information was provided for half the categories; hence, the presence or absence of this information was completely under experimental control, unlike the prior knowledge that subjects bring to an experiment.

There were three specific hypotheses for this experiment, following from the idea that arguments could be treated as equivalent in influence to a fixed number of observations. First, it was predicted that the effect of arguments on categorization should be independent of the effect of observations, just as Heit (1994) found that the effect of prior knowledge was independent of the effect of observations. Second, it was predicted that the arguments would reduce the effects of prior knowledge overall, relative to the standard condition without arguments. Third, it was expected that the integration model would fit the results well, representing the impact of the arguments as being equivalent to some number of observations.

Method

Overview. The subjects observed 160 descriptions of people in City S, a fictional city in England. The descriptions were quite simple; for example, somebody might be described with the terms *shy* and *does not attend parties often*. The descriptions were presented individually. In effect, the subjects were learning about contextualized categories, such as shy people in City S. Following the study phase, the subjects made transfer judgments on additional descriptions of

Table 1
Feature Couplets

shy / not shy
does not attend parties often / attends parties often
more traffic accidents than average / fewer traffic accidents than average
higher car insurance rate than average / lower car insurance rate than average
attends football matches regularly / does not attend football matches regularly
buys football team clothing / does not buy football team clothing
frequently travels by train / does not travel by train often
owns a railway season ticket / does not own a railway season ticket
body builder as a hobby / not a body builder
very muscular / not very muscular
attended a college course in hair and beauty / did not attend a college course in hair and beauty
works as hairdresser / does not work as hairdresser
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

people in City S. For example, the subjects were given the description of another person in City S who does not attend parties often, and they had to judge the likelihood that this person would fall into the *shy* category. The subjects were presented with verbal arguments for half the couplets. For example, some subjects were presented the argument that shy people might actually attend quite a lot of parties, as a means of overcoming their shyness. These arguments were presented before the category members themselves. The subjects read one argument at a time, then rated its convincingness on a numerical scale. The main purpose of collecting ratings was to ensure that the subjects would read the arguments and think about them.

It is best to think of the experiment in terms of the design of the test phase, which had three independent variables. First, half the questions involved the pairing of a description with a category that was congruent according to prior knowledge, such as *does not attend parties often,* and *shy*, and half the questions involved an incongruent pairing, such as *owns a railway season ticket* and *does not travel by train often*. The second variable was presentation frequency, with five possible levels: 12.5%, 25%, 50%, 75%, and 87.5%. These values refer to the proportion of times, out of eight training examples with the description, that the example fell into the category. For example, there might have been eight training examples with the *does not attend parties often* description; of these, two may have been listed as *shy*, and six may have been listed as *not shy*. Hence, the presentation frequency in this case would be 25%. The third variable was whether the subject was provided an argument, relevant to these stimuli.

Subjects. Sixty-four University of Warwick students participated for course credit or a small payment.

Stimuli. The training examples were derived from a set of descriptive terms, shown in Table 1. In total, there were 10 couplets of four features. In the table, each couplet of four features is composed of two pairs of opposites or complements. The stimuli were pretested on other University of Warwick students to validate this manipulation of congruence with prior knowledge (as in Heit, 1994, Appendix A).[1] A verbal argument was prepared for each of the 10 couplets in Table 1. The aim of the argument was to explain why various incongruent features might actually go together. Four sample arguments, out of the total set of 10, are shown in Table 2.

Table 2
Examples of Arguments

Despite the common conception that shy people do not attend parties often, and that un-shy people frequently attend parties, there are many people who do not fall into either of these categories. Many shy people are only shy around strangers; thus they might frequently attend parties of their friends. Shy people may also go to parties in an attempt to overcome their shyness and to meet new people. Conversely, people who are not shy do not always attend parties often. Such people generally have many interests and hobbies, and may be too busy to attend parties often. Also, having an outgoing nature does not necessarily mean that a person will like parties; they may prefer other forms of entertainment.

Although you might think that people who have high car insurance rates would also have more traffic accidents than average, this may not always be the case. Because they have higher car insurance rates, many people become more careful when driving. Additionally, certain groups of people, such as teenagers, are automatically given high insurance rates, even though they may not ever have a car accident. Similarly, people who have lower car insurance rates do not always have fewer traffic accidents than average, because they might become less careful when driving.

Contrary to popular belief, people who attend football matches regularly do not always buy football team clothing. Football team clothing is expensive, and many people may prefer to spend their money on attending an actual game. Also, many people go to small local matches, and so do not buy the clothing associated with big-name football teams. People who do not attend football matches regularly do, however, often buy football team clothing. Often, people who are interested in football are unable to attend actual games because of constraints such as time, money, and location of matches. They may instead show their support of their favorite teams by buying football team clothing.

Many people who travel often by train do so as part of their job, and thus do not own a railway season ticket because their company is paying their travelling costs. Others travel frequently by train, but not frequently enough to make it worthwhile to purchase a season ticket. Conversely, many people who do not travel frequently by train do actually possess a railway season ticket. Many people purchase a season ticket and then, due to changes in circumstances, later may use the railways less frequently. Other people find it cost-effective to purchase a season ticket, even though they do not use the railways very often.

For each subject, the 10 couplets were assigned randomly to the following schema. Each descriptive term appeared in eight training examples, with each example consisting of two pieces of information. For half the couplets, an argument was presented; for the remaining half, there was no argument. Two couplets, one from the argument condition and one from the standard condition, were assigned to each of the following structures. The pairings of characteristics were 12.5%, 25%, 50%, 75%, and 87.5% congruent with prior knowledge, corresponding to the fractions $\frac{1}{8}$, $\frac{2}{8}$, $\frac{4}{8}$, $\frac{5}{8}$, and $\frac{7}{8}$. For example, when the shyness/parties couplet was assigned to the 25% congruent condition, the subjects saw two examples with {*shy, does not attend parties often*} and six examples with {*shy, attends parties often*}. Also, the subject saw two examples with {*not shy, attends parties often*} and six examples with {*not shy, does not attend parties often*}, so that each descriptive term appeared eight times in total.

Each test question in the transfer phase was a conditional probability judgment, 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. The second variable was the conditional probability of presentation during the study phase: 12.5%, 25%, 50%, 75%, or 87.5%. The third experimental variable was whether arguments had been presented. Eight test questions were derived from each of the 10 couplets; thus, there were 80 test questions.

Procedure. All information was displayed on a computer screen. The procedure consisted of four parts. First, the subjects were given the opportunity to familiarize themselves with the stimuli. The features in Table 1 were shown together for 3 min, in a different random order for each subject. The subjects were simply instructed to look over the stimuli.

Second, the subjects read five arguments, rating each one for convincingness. The instructions were as follows:

> After seeing all this descriptive information, you probably have some expectations about how these traits may go together. Just as you might expect that working as a professional engineer and reading science fiction novels would tend to go together, you might believe certain pairs of the facts that you just read would tend to go together. We have asked other people for their opinions, and we would like to know whether you agree with them.

The verbal arguments were presented one at a time, for the subjects to read at their own pace. After each argument was presented, the subjects rated it on a scale from 1 (*not at all convincing*) to 7 (*very convincing*).

Next, at the start of the training phase, the subjects were told that they would see a number of descriptions of persons living in City S, a city located in England. In the training phase, each subject saw the 160 person descriptions, each containing two features, displayed in a random order, one at a time for 3.5 sec. There was a brief gap between displays (0.2 sec), during which the computer screen was cleared. The training phase was followed by a 1-min break.

Finally, in the test phase, the 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 S with the following characteristic: shy
>
> How likely is it that this person would also have this characteristic? attends parties often.

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 S after having seen descriptions of some of the citizens of City S.

The training phase used an unsupervised learning procedure in which the subjects simply observed cases and learned without further feedback. In the test phase, characteristics were treated (in the experimental design) as both predicting features and category labels. For example, the subjects were given the *shy* feature and asked to judge the likelihood of the *attends parties often* category and were given the *attends parties often* feature and asked to judge the likelihood of the *shy* category (see also Heit, 2001, for applications of the integration model to more conventional designs in which there is a sharper distinction between features and category labels and in which subjects classify based on multiple features).

Results

Argument ratings. The subjects found the arguments to be moderately convincing, with an overall mean rating of 4.43 on a 7-point scale. Although this number may seem a bit low, it should be pointed out that the arguments all contradicted people's prior beliefs about stimuli that had been chosen to elicit strong beliefs. The mean ratings for individual arguments ranged from 3.77 to 5.11.

Probability estimates. The results of Experiment 1, in terms of the average probability estimate for each type of test question, are shown in Figure 1, separately for the standard condition (without arguments) and the argument condition. 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 does not attend parties often being shy. 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.

There are a few observations to be made from Figure 1. First, in the standard condition, the parallel-lines pattern is similar to the results of comparable experiments in Heit (1994, 1998), showing independent influences of observations and prior knowledge. Second, in the argument condition, the congruent and incongruent lines are still parallel, but closer together, suggesting that the effect of prior knowledge had been comparatively reduced. Finally, although the arguments did serve to reduce the prior knowledge effect, there was still a substantial difference between the congruent condition and the incongruent condition.

An analysis of variance (ANOVA) supported these observations. There was a main effect of congruent versus incongruent test question [$F(1,62) = 76.85$, $MS_e = 1{,}723$, $p < .001$], with higher probability judgments overall for congruent questions. Also, there was a main effect of presentation frequency [$F(4,248) = 18.71$, $MS_e = 245$, $p < .001$], with higher probability judgments when presented proportions were higher. There was no significant effect of condition (standard or argument) [$F(1,74) = 0.31$, $MS_e = 76$], suggesting that the average probability estimates were about the same for the two conditions. Most critically, the interaction between congruence and condition was significant [$F(1, 62) = 16.76$, $MS_e = 244$, $p < .001$], fitting with the observation that the effect of theory congruence was reduced in the argument condition relative to the standard condition. There was no significant two-way interaction

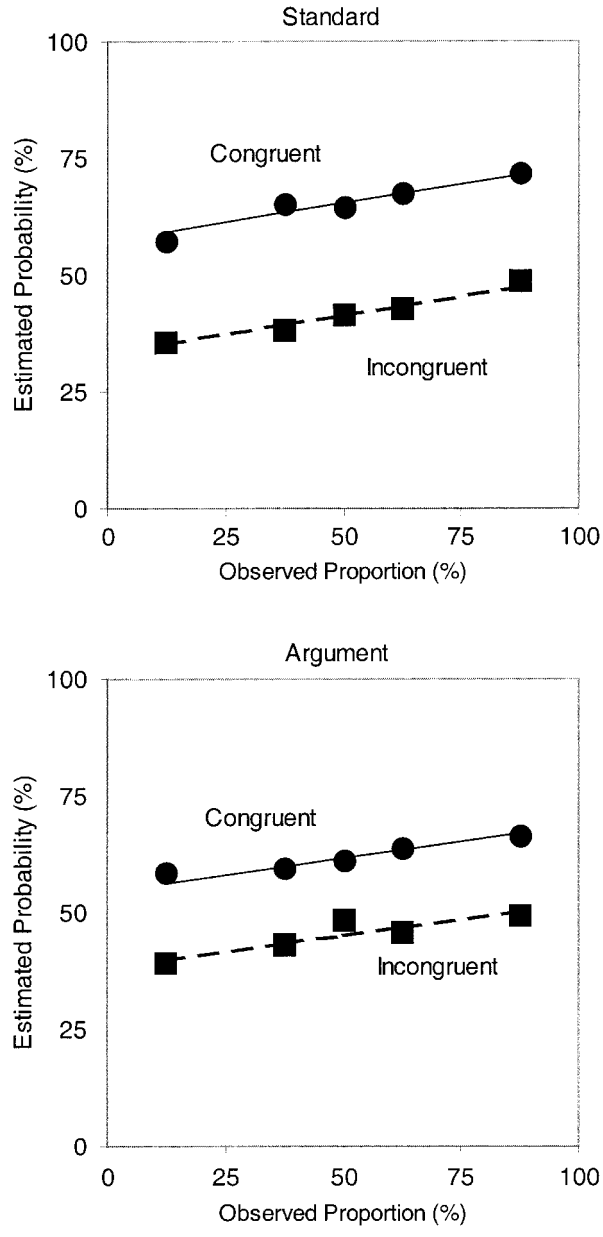

Figure 1. Results of Experiment 1 for the standard and argument conditions and predictions of the integration model.

between presentation frequency and the other two variables, and, likewise, the three-way interaction was not statistically significant.

Model-based analyses. Next, the results were analyzed in terms of a categorization model. Unlike the ANOVA, which looked at individuals' responses, the model was applied to average responses at the group level. In particular, the integration model of Heit (1994) was applied to the results, taking into account the subjects' prior knowledge, the observed category members, and the arguments when presented. The model is presented in detail in the Appendix. What is most critical is that the integration model has a free parameter, *G*, which represents the strength of prior knowledge. Prior knowledge and observations are simply added together to make a judgment, so that the numerical value of *G* can be interpreted as being equivalent to some number of observations. For example, if *G* is 5, then prior knowledge has the equivalent impact of 5 observations. The categorization model also has another free parameter, *s*, indicating the degree of memory confusions (Medin & Schaffer, 1978). The value of *s* is in the range from 0 to 1, with higher values of *s* indicating worse memory.[2]

The integration model was applied to the data from both the standard condition and the argument condition together. It was assumed that, in the standard condition, the subjects would be influenced by their prior knowledge, reflected by the *G* parameter, as well as by the observed category members. In the argument condition, the subjects would be influenced by prior knowledge, represented by the same value of *G*, by observed category members, and by the argument, with a strength given by the parameter *T* (see the Appendix for further details). The *G* parameter, representing strength of prior knowledge, was estimated to be 11.9. The *T* parameter, representing the strength of the argument, was estimated to be 2.6. Finally, the *s* parameter was estimated to be .43.

The root mean square error (*RMSE*) of this model fit was quite low, .0123. The predictions of the integration model are shown as the lines in Figure 1, superimposed over the data points. Notably, the model captures the difference between congruent and incongruent test questions in the standard condition, and it also reflects the reduced effect of prior knowledge in the argument condition.

Discussion

The results supported all the experimental hypotheses. That is, both verbal arguments and prior knowledge had opposing effects on categorization, not interacting with presentation frequency. Relative to the standard condition, the effects of prior knowledge were reduced in the argument condition, in an overall manner predicted by the integration model. These results give converging evidence for the theoretical account developed in Heit (1994, 1998), and, perhaps most important, these results extend the domain of categorization models to account for effects of verbal arguments.

The finding that arguments can be treated as equivalent to a fixed number of observations, with an additive influence on judgments, is informative about the role of arguments in reasoning. According to the integration account, the arguments have a fixed impact on people's beliefs, reducing the prior knowledge effect uniformly in Figure 1. Another possibility would be that arguments simply have a cuing effect, not actually changing subjects' beliefs but merely letting them know that some observations may be different from what is expected, facilitating the use of this information. If this were the case, then arguments might be expected to have a greater impact in some conditions than in others, (e.g., arguments should have a greater impact when the observations are 87.5% incongruent than when the observations are only 12.5% incongruent). However, again, the fixed effect of arguments in Figure 1 rules out this explanation.

Interestingly, the effect of prior knowledge was fairly large in this experiment, relative to previous experiments with similar designs. The strength of the prior knowledge effect in the standard condition was estimated to be equivalent to 11.9 prior examples. The exact value of *G* should not be taken too literally in comparing experiments that had somewhat different designs and that were run on different subjects, but the value of 11.9 is higher than that estimated in any comparable previous experiment in this series. Compared with Heit (1994, Experiments 1–4), Heit (1995), Heit (1998, Experiments 1–3), and several unpublished pilot experiments, the effect of prior knowledge in the present experiment was the highest—past estimates of *G* have been in the range of 1 to 7. However, in a replication of Experiment 1, using a similar procedure including the presentation of arguments, the estimated value of *G* was approximately 8. So there is not strong evidence that presenting arguments has a quantitative effect on the strength of prior knowledge effects

EXPERIMENT 2

Heit (1998) found that when training occurred at a slower pace, with more than 10 sec per observation, there was an interaction between the two variables of interest, with the effect of prior knowledge being reduced for mixed observations (near the 50% range) relatilve to unmixed observations (near the 0% and 100% range). The parallel-lines pattern illustrated by Figure 1 is replaced by two curved lines, closer to each other in the 50% region. Applying models to this curved pattern suggested a greater influence of incongruent observations than of congruent observations. Hence, it was concluded that under slower learning conditions, prior knowledge has two distinct influences. First, there is an integration effect such that prior knowledge has an influence equivalent to a fixed number of observations. Second, there is also a selective weighting effect, favoring incongruent category members.

The main purpose of Experiment 2 was to provide converging evidence for this account, applying it to verbally presented arguments. In Experiment 1, the study times were brief (less than 4 sec per category member), whereas in Experiment 2, the study time was 16 sec for each cat-

egory member. In Experiment 1, the efficacy of verbal arguments in reducing prior knowledge effects helped to support the idea that under fast learning conditions, prior knowledge has a simple main effect on categorization judgments. In Experiment 2, the question of interest was whether arguments could reduce both prior knowledge effects that have been observed with slower paced learning procedures, integration and selective weighting. It was expected that the arguments would act like prior knowledge and, hence, also have two influences, but working in the opposite direction to the prior knowledge. Thus, it was predicted that both influences of prior knowledge would be counteracted by the arguments.

Method

Overview. Experiment 2 was similar to Experiment 1, in that the subjects were presented with arguments before they observed descriptions of residents of City S. The main change was category members were presented in the manner of Heit (1998): The subjects saw a smaller number of person descriptions (40), with more information in each description (a name and four other characteristics), and these descriptions were presented at a much slower pace (16 sec per item). Heit (1998, Experiment 2) found that study time was the key variable for encouraging optional selective weighting processes; the other changes in the materials were not sufficient to lead to selective weighting. However, the number and the length of person descriptions were also patterned after Heit (1998) to maximize the chance of finding selective weighting.

Subjects. Fifty-three University of Warwick students participated for course credit or a small payment.

Stimuli. Each training example was a description of a person in terms of a name and four characteristics. For example, someone was named H Eccles, was shy, did not attend parties often, traveled frequently by train, and owned a railway season ticket. The 40 names (first initial and surname) were chosen at random from the Coventry, Nuneaton, and Rugby (England) telephone directory.

The 40 training examples were constructed from the features in Table 1. Each example contained two pairs of features, chosen randomly from two different couplets. An example contained either two congruent pairings or two incongruent pairings. To construct these examples, two couplets of features were assigned to each level of congruency, 0%, 25%, 50%, 75%, and 100%, corresponding to the fractions 0/4, 1/4, 2/4, 3/4, and 4/4. Each particular feature (e.g., *shy*) appeared four times. For example, when the shyness/parties couplet was assigned to the 100% condition, there were four examples describing someone who is shy and does not attend parties often, as well as four examples describing someone who is not shy and does attend parties often. Similarly, when the train/season ticket couplet was assigned to the 25% congruent condition, there was one example with {*frequently travels by train*, *owns a railway season ticket*}, one example with {*does not travel by train often*, *does not own a railway season ticket*}, three examples with {*frequently travels by train*, *does not own a railway season ticket*}, and three examples with {*does not travel by train often*, *owns a railway season ticket*}.

For each pair of couplets assigned to the same level of congruency, one was selected to have arguments presented. For example, when the shyness/parties couplet and the traffic accidents/car insurance couplet were both assigned to the 100% congruent condition, one of them (say, the shyness/parties couplet) had an argument presented, and the other did not.

What is most critical is the design of the test stimuli, with three within-subjects factors. Each test question was a conditional probability judgment for a pair of features taken from a couplet in Table 1, 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. 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 was equivalent to the percentage of congruent presentations during the training phase; on test questions involving two incongruent features, the percentage was equivalent to the percentage of incongruent presentations during the training phase. The third variable was whether an argument was presented for this couplet. Eight test questions were derived from each couplet; thus, there were 80 test questions.

Procedure. The procedure was like that of Experiment 1, with five arguments presented to each subject. The main change was in the training phase, in which each subject saw 40 person descriptions, one at a time for 16 sec. The descriptions were presented as in the following example:

H Eccles has this description:

shy

does not attend parties often

frequently travels by train

owns a railway season ticket

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*) were kept adjacent to each other.

Results

Argument ratings. The subjects found the arguments to be moderately convincing, with an overall mean rating of 4.26 on a 7-point scale. The mean ratings for individual arguments ranged from 3.23 to 5.00.

Probability estimates. The results, in terms of the average probability estimate for each type of test question, are shown in Figure 2, separately for the standard condition and the argument condition. There are a few observations to be made. First, in the standard condition, the curved pattern of lines resembles that of Heit (1998, Figure 2), showing the characteristic pattern of interaction when incongruent category members have more influence than congruent category members. In the argument condition, congruent and incongruent lines have moved closer together, on average, but, more strikingly, the interactive pattern of results has been replaced by the parallel-lines pattern. The results in the argument condition have the characteristic pattern when congruent and incongruent category members have equal influence.

An ANOVA supported these observations. There was a main effect of congruent versus incongruent test question [$F(1,52) = 22.71$, $MS_e = 1{,}353$, $p < .001$], with higher probability judgments overall for congruent questions. Also, there was a main effect of presentation frequency [$F(4,208) = 91.83$, $MS_e = 382$, $p < .001$], with higher probability judgments when presented proportions were higher. There was no significant effect of condition (standard or argument) [$F(1,74) = 0.07$, $MS_e = 68$], suggesting that the average probability estimates were about the same for the two conditions. More critically, the interaction between congruence and condition was significant [$F(1,52) =$

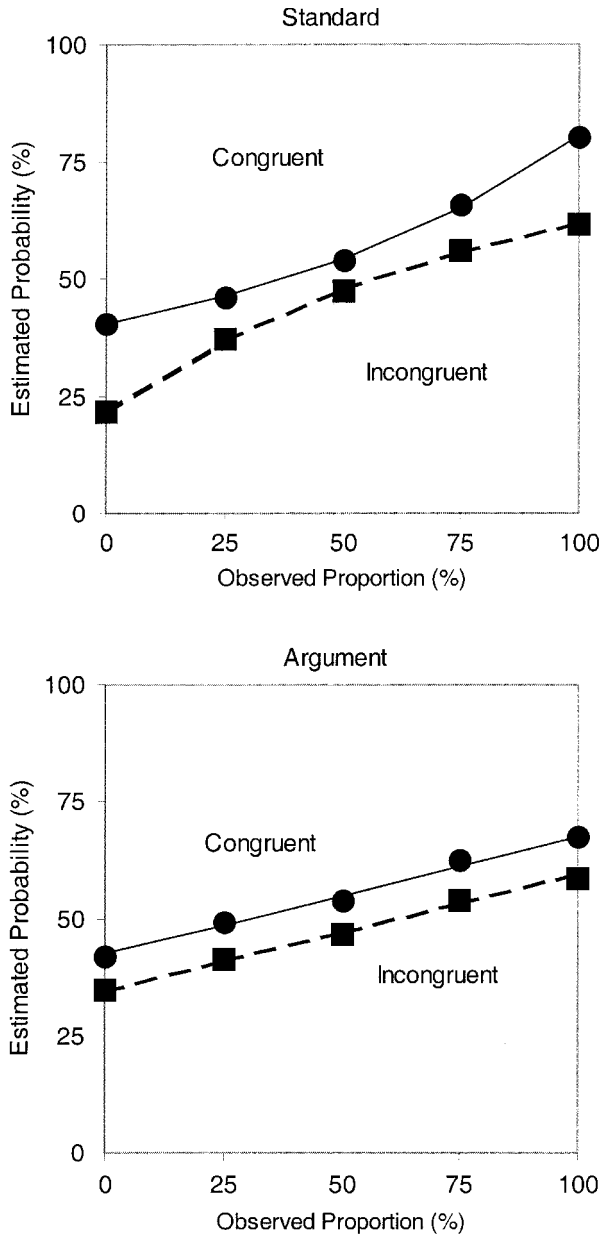

Figure 2. Results of Experiment 2 for the standard and argument conditions and predictions of the integration model with selective weighting in favor of incongruent category members.

4.33, MS_e = 362, $p < .05$], fitting with the observation that the fixed effect of theory congruence was reduced in the argument condition relative to the standard condition. Also, there was an interaction between condition and presentation frequency [$F(4,208) = 11.58$, $MS_e = 155$, $p < .001$], suggesting that the subjects were differently sensitive to the observed category members in the two conditions. The final two analyses addressed the differential effect of prior knowledge at different levels of presentation frequency, which is relevant to selective weighting. The interaction between congruence and presentation frequency did not quite reach statistical significance [$F(4,208) = 2.07$, $MS_e = 258$, $p < .09$], but this result should be interpreted in light of the three-way interaction, between these two variables and condition, which is more relevant. This three-way interaction reached the level of statistical significance [$F(4,208) = 2.91$, $MS_e = 161$, $p < .05$], suggesting that the interaction between congruence and presentation frequency was different in the standard and argument conditions.

For a further look at this interaction, data from the standard and argument conditions were examined separately, with additional analyses previously used in Heit (1998). For each condition, 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 each panel in Figure 2, 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. For the standard condition, trend analyses revealed that the difference scores showed a significant quadratic component as a function of level of presentation [$F(1,208) = 17.50$, $MS_e = 224$, $p < .001$]. 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 approached the middle of the range (50%). For the argument condition, trend analyses did not indicate a significant quadratic trend [$F(1,208) = 0.02$, $MS_e = 195$], fitting with the visual pattern in Figure 2 suggesting that there no interaction in the argument condition. These analyses provide strong evidence at the level of individual subjects' responses for the selective weighting pattern in the standard condition but not in the argument condition.

Model-based analyses. The integration model was applied to the average responses together from the standard and argument conditions, as in Experiment 1. The main change was that some degree of selective weighting of category members was allowed. The integration model with an added weighting component is described in the Appendix. Following Heit (1998), it was expected that, with slower paced learning, there would be a greater influence of incongruent category members than of congruent category members. The relative weight of incongruent category members was represented by the *W* parameter. For example, if *W* is 2, then incongruent category members have twice the influence of congruent category members on categorization. If *W* is 1, then both kinds of category members have the same influence, and the integration plus weighting model is equivalent to the original integration model.

For the initial model application, separate *W* parameters were estimated to measure the degree of weighting in the standard and argument conditions. The *G* parameter was used to measure the strength of prior knowledge, and the *T* parameter was used to measure the strength of the argument, both in terms of an equivalent number of obser-

vations. In addition, a single value of the *s* parameter, representing memory confusions, was estimated for both conditions, as in the previous experiment. This model application was fairly successful, with an *RMSE* of .0181. However, it was clear that there was a systematic discrepancy between the data points and the model's predictions. Namely, relative to the model's predictions, the subjects' estimates were more sensitive to observed proportion in the standard condition and were less sensitive to observed proportion in the argument condition. That is, the model seemed to get the slope of the lines somewhat wrong, relative to the data points in Figure 2. This finding suggested that the subjects were differentially sensitive to observed category members in the two conditions, a result also suggested by the ANOVA, particularly the interaction between condition and presentation frequency.

In an effort to capture this trend, the model was fitted again with two separate *s* parameters, allowing for different degrees of category member confusions in the two conditions. The model fit was improved, with an *RMSE* of only .0059. This improvement for the model with separate estimates of memory confusions in the two conditions was statistically significant, beyond what would be expected by simply adding another free parameter, [$\chi^2(1) = 36.08$, $p < .001$].[3] (Note that this change did not lead to significant improvements for Experiment 1.) The estimated parameter values were as follows. The degree of weighting, *W*, in the standard condition was 3.15, and the degree of weighting in the argument condition was 1.05. That is, in the standard condition, incongruent category members had about three times the influence of congruent category members, but, in the argument condition, the relative influences were about equal. The strength of prior knowledge, *G*, was estimated to be 1.87, and the strength of the argument, *T*, was estimated to be 0.55. The *s* parameter for the standard condition was .26, and the *s* value for the argument condition was .43, suggesting a greater degree of category member confusions in the argument condition. The predictions for this model are shown as lines in Figure 2, overlaid on the data points. The correspondence between model and data is excellent.

Additional analyses were conducted to establish the level of statistical significance for the estimates of *W* (see Heit, 1998). Two additional, restricted models were fit to the data. For one model, the degree of weighting for the standard condition was fixed at 1 (i.e., there was no selective weighting). This model had an *RMSE* of .0194, which was significantly worse than the model that allowed for selective weighting in the standard condition [$\chi^2(1) = 38.32$, $p < .001$]. Hence, it appears that selective weighting made a statistically significant contribution in the standard condition. Next, a model was applied that fixed the degree of weighting in the argument condition to 1. This model had an *RMSE* of .0060, which was not significantly worse than the model that allowed for selective weighting in the argument condition [$\chi^2(1) = 0.77$]. Hence there was no evidence for a weighting component in the argument condition. (Likewise, using comparable analyses, there was no evidence at all for selective weighting in Experiment 1.)

Discussion

The results fit well with predictions and with past findings. As in previous experiments (Heit, 1998) with slower paced learning procedures, in the standard condition the modeling revealed two effects of prior knowledge: a fixed effect due to initial estimates, and selective weighting of incongruent category members. Thus, the standard condition was a useful replication of previous results. These findings were also supported at the individual-subject level by the ANOVA and trend analysis.

Also, as predicted, the arguments acted like prior knowledge, but working in the opposite direction. That is, the arguments reduced the fixed effect of prior knowledge represented by the *G* parameter, by providing a number *T* of counterexamples working in the opposite direction. Likewise, the arguments eliminated the selective weighting of incongruent category members represented by the *W* parameter—in the argument condition, it was estimated that there was no selective weighting of category members at all. These results provide converging evidence for the theoretical account developed in Heit (1994, 1998), suggesting that there are indeed two distinct influences of prior knowledge under slower learning conditions.

Although it had not been expected that modeling the results would require different values of the *s* parameter, representing memory confusions, for the standard and argument conditions, this result is easy to accommodate in terms of the arguments leading to worse overall learning of related category members, relative to category members, that did not have arguments presented. There was not significant evidence for this difference in Experiment 1, but, possibly, Experiment 2 was more sensitive to variation in memory for observations due to the greater time of presentation per item. Furthermore, the issue of memory confusions is separate from the other issues addressed by the analyses and does not detract from the overall qualitative pattern of results. The other differences between the standard and argument conditions in Experiment 2 are more important—namely, that arguments led to a reduction in both influences of prior knowledge, a fixed effect equivalent to some number of observations as well as an effect of selective weighting. In sum, the within-experiment comparisons clearly showed that prior knowledge itself had two distinct effects, and arguments acted like prior knowledge but worked in the opposite direction, serving to reduce both of these effects.

GENERAL DISCUSSION

These experiments provided converging evidence for the theoretical account of prior knowledge effects developed in Heit (1994, 1998). Under faster paced learning conditions in Experiment 1, verbal arguments had a fixed influence on categorization, equivalent to some number of observed category members. Under slower learning conditions in

Experiment 2, verbal arguments also had a second effect, affecting the selective weighting of category members.

Much of what we learn about categories is learned not by direct observation of category members but is conveyed through language, such as in written explanations, conversations, and arguments. Until now, this kind of information has also been out of the bounds of categorization models, a situation bearing some similarity to the case of prior knowledge also being excluded from most categorization modeling efforts. Indeed, it was observed that verbal arguments acted much like the subjects' own prior knowledge in terms of qualitative influences. These results support the hypothesis that verbally presented arguments can be treated in the same way as other forms of prior knowledge, from the perspective of applying categorization models.

The present modeling efforts revealed a few novel results regarding the influences of verbal arguments. First, in Experiment 1, it was observed that arguments changed people's beliefs (as well as could be observed from their judgments) in a similar way to observations. That is, the arguments did not simply prepare people for contradictory evidence, but they actually served as contradictory evidence. Second, in Experiment 2, it was possible to determine whether the arguments act more like observations or more like a person's prior knowledge, because, in this experiment, these two sources of information had different qualitative effects. It was discovered that verbal arguments actually acted more like people's own knowledge. Verbal arguments not only changed beliefs, but they also affected the way that people treated new observations of category members. This result could possibly have practical implications—for example, presenting an argument in an advertisement could be better than simply showing a happy customer, because the argument could have persistent effects on how future information is used.

Perhaps what these experiments illustrate at the most general level is that the borders of categorization modeling can be expanded from previous conceptions. Rather than designing categorization experiments that avoid effects of subjects' previous knowledge, it is possible to design experiments that profit from the consistent qualitative patterns of prior knowledge effects. Furthermore, the issue of how to incorporate verbal arguments into categorization can be addressed in this same experimental paradigm and with the same models. Clearly there is more work to be done. For example, recently Heit and Bott (2000) posed the *knowledge selection* problem, which points out that although prior knowledge might facilitate people's learning, there is still the potentially difficult problem of figuring out which prior knowledge will be helpful for learning from among many possible sources of prior knowledge. Despite the difficulties of this problem, Heit and Bott did present a computational model that addresses some aspects of knowledge selection. The present experiments, showing that arguments can be treated, systematically, as yet another source of knowledge guiding category learning reinforce the idea that accommodating multiple sources of prior knowledge is central to categorization.

REFERENCES

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

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

Hastie, R. (1980). Memory for behavioral information that confirms or contradicts a personality impression. In R. M. Hastie, T. M. Ostrom, E. B. Ebbesen, R. S. Wyer, D. L. Hamilton, & D. Carlston (Eds.), *Person memory: The cognitive basis of social perception* (pp. 155-178). Hillsdale, NJ: Erlbaum.

Hayes, B., Taplin, J., & Munro, K. (1996). Prior knowledge and sensitivity to feature correlations in category acquisition. *Australian Journal of Psychology,* **48**, 27-34.

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

Heit, E. (1995). Belief revision in models of category learning. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth 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. (1998). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **24**, 712-731.

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

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

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

Macrae, C. N., Bodenhausen, G. V., Schloerscheidt, A. M., & Milne, A. B. (1999). Tales of the unexpected: Executive function and person perception. *Journal of Personality & Social Psychology,* **76**, 200-213.

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.

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

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

Rotello, C., & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory & Language,* **40**, 432-453.

Spalding, T. L., & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **22**, 525-538.

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology,* **28**, 274-328.

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

NOTES

1. In the pretest with 24 additional subjects, the average probability estimate for incongruent pairings in Table 1 (e.g., *shy* and *attends parties often*) was 25%, and the average estimate for congruent pairings (e.g., *generous* and *donates to charity*) was 75%.

2. In addition, the average responses in each experiment were adjusted by a calibration parameter. The subjects in these experiments showed a slight lack of calibration, such that complementary probabilities added to somewhat more than 100% (see also Heit, 1994, 1998).

For example, the average response in Experiment 1 was 52.3%. To compensate for this lack of calibration, a value (here of .023) was subtracted from each response.

3. 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 Model 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, 20). Borowiak showed that $-2 \ln(\lambda)$ has a χ^2 distribution with $(a - b)$ degrees of freedom (see Heit, 1998, and Rotello & Heit, 1999, for other applications of this technique).

APPENDIX
Description of Categorization Models

The integration model of categorization is an exemplar-based account that extends the model of Medin and Schaffer (1978). The main difference is that the representation of a category includes not only observed category members but also includes prior knowledge in the form of prior examples. Exemplar models are based on Equation A1, which describes classifying a stimulus, x, as a member of either Category A or Category B

$$P(\text{classify } x \text{ as A}) = \frac{\text{fam}_A(x)}{\text{fam}_A(x) + \text{fam}_B(x)}. \quad (1)$$

Categorization depends on the underlying psychological construct of familiarity. The familiarity of stimulus x is evaluated with respect to the members of the two categories. The likelihood of categorizing x as an A increases with fam_A, the familiarity of x with respect to A, and decreases with fam_B, the familiarity of x with respect to B. In Equation 2A, the familiarity of stimulus x with respect to Category A is the sum of the similarities of x to each member of A retrieved from memory. Likewise, in Equation 2B, $\text{fam}_B(x)$ is the sum of similarities of x to each member of B retrieved from memory. Here, the similarity between two identical stimuli, $\text{sim}(x,x)$, is assigned the value of 1, and the similarity between x and a mismatching stimulus y, $\text{sim}(x,y)$, is assigned a fractional value s, where $0 \le s < 1$. This implementation of the sim function is a special case of the multiplicative similarity rule of Medin and Schaffer (1978) for stimuli described by a single feature. The free parameter s measures the ability of a subject to discriminate between stimuli retrieved from memory. When s is 0, memory discrimination is perfect and mismatching stimuli have no effect on categorization.

$$\text{fam}_A(x) = \sum_{a \in A} \text{sim}(x,a) = \text{No. } x \text{ in A} + s(\text{No. } y \text{ in A}) \quad (2A)$$

$$\text{fam}_B(x) = \sum_{b \in B} \text{sim}(x,b) = \text{No. } x \text{ in B} + s(\text{No. } y \text{ in B}) \quad (2B)$$

Finally, Equations 2A and 2B may be substituted into Equation 1, leading to Equation 3, which is the basic model applied in this paper:

$$P(\text{classify } x \text{ as A}) = \frac{\text{No. } x \text{ in A} + s(\text{No. } y \text{ in A})}{\text{No. } x \text{ in A} + s(\text{No. } y \text{ in A}) + \text{No. } x \text{ in B} + s(\text{No. } y \text{ in B})}. \quad (3)$$

In fitting this model to Experiments 1 and 2, it was assumed that an additional number, G, of prior examples would be represented in each category. Consider judgments when description x is expected to be in Category A and description y is expected to be in Category B. Here, for the purpose of applying Equation 3, the number of x stimuli in Category A would be the observed number plus G, and, likewise, the number of y stimuli would be the observed number plus G. For example, if a subject observes 7 shy people who avoid parties, and G is estimated to be 5, then, for Equation 3, the number of shy people who avoid parties would be 12. The value of G was estimated as a free parameter, using a least squares criterion (see Heit, 1994, for further details).

Also, it was assumed in the argument condition that some number, T, of counterexamples would also affect judgments. These counterexamples would work against prior knowledge; so, in the present case, there would be T examples of x in Category B and, likewise, T examples of y in Category A also included in Equation 3. The value of T for the argument condition was estimated as a free parameter.

For Experiment 2, there was an additional extension. Here, in counting up the numbers of observed category members in the application of Equation 3, different category members had different influences. Theory-congruent category members, such as generous people who give to charity, were always given an influence of 1. However, incongruent category members, such as generous people who do not give to charity, were given an influence of W, where W was an estimated free parameter. When W is estimated to be greater than 1, the data indicate than incongruent category members have a greater influence than do congruent category members. For an illustration, say that a subject sees 4 generous people who give to charity, and 4 generous people who do not give to charity, and W is estimated to be 3. For the purpose of putting numbers of observed category members into Equation 3, the observed number of generous people who give to charity would be 4 and the observed number of generous people who do not give to charity would be 12 (see Heit, 1998, for further details).

(Manuscript received June 27, 2000;
revision accepted for publication May 25, 2001.)