Journal of Experimental Psychology: General
2005, Vol. 134, No. 4, 596–605

0096-3445/05/$12.00 DOI: 10.1037/0096-3445.134.4.596

Relations Among Categorization, Induction, Recognition, and Similarity: Comment on Sloutsky and Fisher (2004)

Evan Heit
University of Warwick

Brett K. Hayes
University of New South Wales

V. M. Sloutsky and A. V. Fisher (2004) reported 5 experiments documenting relations among categorization, induction, recognition, and similarity in children as well as adults and proposed a new model of induction, SINC (similarity, induction, categorization). Those authors concluded that induction depends on perceptual similarity rather than conceptual knowledge. Despite the useful contributions of this work, there are some important limitations. The experimental designs examined a limited range of phenomena that are not the most revealing about the use of nonperceptual information. The main results involved a simple triad task, for which the SINC model's predictions are equivalent to the predictions of previous models of inductive reasoning. It is also unclear whether the SINC model can account for the observed relations between similarity and recognition. Implications for future work on induction and related cognitive activities are discussed.

Keywords: categorization, inductive reasoning, similarity, recognition memory, cognitive development

Three crucial and closely related cognitive activities are categorization, induction, and recognition, which allow people to draw inferences based on their observations and their previous knowledge. For example, categorization allows people to link an individual item to a coherent reference class, such as categorizing some animal as a cow. Induction allows people to infer further properties based on category membership, such as that a cow will have a heart. Recognition allows people to judge whether an individual has been observed before, such as whether this particular cow has been seen before. Indeed, these three activities are so closely related that the defining line among them is sometimes obscure. The boundary between properties and categories can be blurred, so that having a heart can be considered a property but animals with hearts can be considered a category. Hence, whether a particular inference should be called induction or categorization may sometimes be unclear (Anderson, 1991, but see Markman, 1989). In addition, recognition judgments themselves can be considered as a kind of categorization judgment, in which an individual is assigned either to the category of things previously observed or to the category of novel items (Estes, 1994).

The potentially unclear boundaries among categorization, induction, and recognition should be seen not as a problem but rather as an opportunity. That is, the close relations among these activities invite a common psychological account that addresses all three. Curiously, most models of categorization, inductive reasoning, and recognition memory address just one of these activities and not two or three. For example, models of category-based inductive reasoning (Heit, 1998; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975; Sloman, 1993) do not address any phenomena in categorization. The main exception to this point would be descendents of Medin and Schaffer's (1978) context model of categorization, which have been applied not only to categorization but also to induction and recognition (Estes, 1994; Heit, 1993; Lamberts, Brockdorff, & Heit, 2003; Nosofsky, 1988). These exemplar models follow the total similarity principle (Jones & Heit, 1993), namely, that the tendency to make a positive response to some stimulus item is a positive function of the total similarity between that stimulus and other items belonging to a target class compared with its similarity to an alternative class. For example, whether an item is judged as belonging to some category depends on its similarity to other known category members relative to its similarity to nonmembers. Likewise, whether an item is judged to have some property would depend on its similarity to other entities known to have that property. Finally, whether an item is recognized as "old" depends on its similarity to other old, previously observed items. Therefore, similarity is potentially a common currency relating all three of these cognitive activities. Still, it is important not to oversimplify matters. Within categorization and induction, it appears that other forms of knowledge such as theories and explanations have an influence beyond similarity for both adults (Heit, 2001; Medin, Coley, Storms, & Hayes, 2003; Murphy & Medin, 1985; Rehder & Hastie, 2001, 2004) and children (Hayes, Foster, & Gadd, 2003; Krascum & Andrews, 1998). Recognition itself may be influenced not only by similarity but also by recollection (Rotello & Heit, 1999; Yonelinas, 2002).

A recent article by Sloutsky and Fisher (2004) made an important and original contribution to this area of research with a combination of experimentation on children and adults and mathematical modeling. The main issue addressed was how similarity is used in both categorization and induction and how this use

Evan Heit, Department of Psychology, University of Warwick, Coventry, United Kingdom; Brett K. Hayes, School of Psychology, University of New South Wales, Sydney, New South Wales, Australia.

This research was supported by Australian Research Council Discovery Grant DP0344436 and a grant from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

Correspondence concerning this article should be addressed to Evan Heit, who is now at the University of California, Merced, P.O. Box 2039, Merced, CA 95344. E-mail: eheit@ucmerced.edu

changes over the course of development. The relation between similarity and categorization in the other direction was also considered, namely whether sharing a category label would make two items seem more similar. In addition, this article addressed how making an inductive judgment about a stimulus would affect its subsequent recognition, on the assumption that induction and recognition would access a common representation.

In this commentary we first summarize the contributions made by Sloutsky and Fisher (2004). Then, we consider the relations between Sloutsky and Fisher's experimental work and previous research on category-based induction. This is followed by an assessment of Sloutsky and Fisher's mathematical model and its relations to previous models. Next, we consider Sloutsky and Fisher's investigation of links between induction and recognition. Finally, we briefly suggest some implications for future work.

Contributions of Sloutsky and Fisher (2004)

One of the important contributions of Sloutsky and Fisher (2004) was to document the role of category labels. In Experiment 1, they showed that 5-year-old children were heavily influenced by category labels when asked to make a judgment about perceptual similarity. Children were more likely to judge two pictures of animals as looking alike when they were given the same name. In comparison, adults' judgments of perceptual similarity were not affected by shared category labels.

Although the similarity results are interesting in their own right, their main importance derives from their systematic relation to categorization, induction, and recognition. Experiments 2, 3, and 4 had similar designs but had somewhat different stimuli. These experiments used a triad task, following Gelman and Markman (1986), in which a child was shown three pictures: the target picture, Test A, and Test B. In terms of pretesting, perceptual similarity was greater between the target picture and Test A than between the target picture and Test B. In the induction condition, the child was told about distinctive properties of A and B, for example, that A has red bones and B has green bones, and then was asked whether the target would have the A property or the B property. In the categorization condition, no properties were given, and the child was asked whether the target was the same kind of animal as A or B. These two conditions were conducted either without any category labels given or with category labels, in which case the target and Test B shared a category label. For example, the target and Test B would each be called a "lolo," and Test A would be called a "tippy."

There were several important findings. First, the results for categorization and induction judgments appeared to be very similar. Sloutsky and Fisher (2004) did not report any differences between these conditions. In Experiment 2, for example, the correlation between induction and categorization results was .97. Second, the patterns of children's induction and categorization judgments were closely related to the pattern of perceptual similarity judgments in Experiment 1. Children were clearly influenced by the perceptual appearance of the pictures, for example, favoring Test A over Test B. In addition, children were influenced by category labels when presented, for example, showing an increased tendency to choose Test B over Test A. Third, the overall pattern of results was well accounted for by Sloutsky and Fisher's mathematical model, SINC, relating similarity, categorization, and induction. (The SINC model is discussed in more detail in subsequent sections of this article.)

Experiment 5 used an innovative method to examine the relations between induction and recognition. Five-year-old children and adults were shown pictures illustrating that a property possessed by a target animal—for example, a cat—was shared by members of the same category—that is, other cats—but not members of other animal categories—such as bears. Note that labels were not stated for these pictures; however, the pictures had been selected on the basis of a pretest to ensure that they could be easily labeled by children. After completing the induction task, the children and adults were given a surprise recognition memory test, in which they were asked to discriminate between pictures used in the induction task and novel pictures of animals. There were two interesting findings. First, recognition performance by children exceeded that of adults. That is, children were able to easily distinguish pictures of cats that had been previously observed from pictures of cats that had not been observed, but adults showed poor discrimination. These results suggest that children encoded perceptual details during induction, whereas adults encoded category-level information; hence, adults did not distinguish between cats they had previously seen and cats they had not seen. Second, children could be trained to perform induction in an adult manner, focusing on categories, so that their recognition performance fell to adult levels.

Overall, the main conclusion from Sloutsky and Fisher (2004) was that children's induction, as well as categorization, depends on perceptual similarity. Perceptual similarity itself could be influenced by shared category labels. However, this explanation based on similarity was sharply contrasted with an explanation based on further assumptions about nonperceptual information and category structure, for example, use of taxonomic knowledge and beliefs that categories have essences (e.g., Gelman, 2003). Sloutsky and Fisher drew the strong conclusion from their studies that "categorization and induction in young children are similarity-based processes" (p. 185) that are "computed over visual and linguistic cues" (p. 184). In sum, it was claimed that categorization and induction depend on visual matches between stimuli as well as auditory matches in terms of spoken category labels.

The Wider Context of Research on Inductive Reasoning

To examine the contribution of Sloutsky and Fisher's (2004) studies more fully, we thought it best to consider the studies' relation to the wider context of research on inductive reasoning. In this section we focus on previous experimental work on induction, and in the following section we address models of induction (see Heit, 2000, and Heit & Hahn, 2001, for more detailed reviews).

Sloutsky and Fisher's (2004) studies of induction were mainly based on a triad task in which children in effect chose the stronger of two inductive arguments, as follows:

$$(1)\ \frac{\text{Test A has Property X}}{\text{Target has Property X}}$$

$$(2)\ \frac{\text{Test B has Property Y}}{\text{Target has Property Y}}$$

There are a few points to make about these arguments. First, each argument has a single premise, either that Test A has a property or that Test B has another property. The target item is referred to in the conclusion of both arguments. Therefore, these studies addressed how people reason from a single premise, but they did not address how people put together information from multiple premises. These arguments allow an assessment of whether Test A or Test B leads to stronger inferences about the target item, but they do not allow a comparison of different conclusions about different targets. Second, it should be noted that the arguments concern individual animals. That is, Test A, Test B, and the target each corresponded to individual animals, such as individual cats or dogs, rather than categories, such as all cats or all animals. The task used in these studies did not explicitly require children to draw inferences about categories but only about individuals. Third, it should be reiterated that the three crucial items—Test A, Test B, and the target item—were each presented as pictures (photographs or drawings). Hence, visual information about these items was readily available.

In comparison to the wider range of research on induction in children and adults, the Sloutsky and Fisher (2004) studies covered only a narrow range of the possibilities. The first narrowing is due to the limitations of the triad design. Although the use of the triad design has some benefits, such as direct comparability with Gelman and Markman's (1986) classic experiments, there is a much wider range of inductive phenomena for children and adults (Heit, 2000). For example, Osherson et al. (1990) described 11 main phenomena in category-based induction, yet only 2 of these could be addressed with the triad design used by Sloutsky and Fisher. One of these phenomena concerns the similarity between the premise and conclusion items. This was the issue actually addressed by Sloutsky and Fisher. The other phenomenon concerns the typicality of the premise item, that is, the extent to which the premise item is similar to known category members. More typical premise items tend to lead to stronger inferences. Even novel stimuli would vary in typicality if they vary in their similarity to known items. The issue of typicality was not considered by Sloutsky and Fisher, although typicality could have also affected the results. Indeed, typicality effects could point to the use of conceptual knowledge in addition to perceptual information. The remaining 9 phenomena would require either presenting multiple premises in each argument or allowing the conclusion to vary between arguments. For example, adults (and, in some cases, children) are sensitive to the diversity of presented evidence, as in the following arguments:

(3)
Hippos require Vitamin K for the liver to function.
Rhinos require Vitamin K for the liver to function.
———————————————
Elephants require Vitamin K for the liver to function.

(4)
Hippos require Vitamin K for the liver to function.
Hamsters require Vitamin K for the liver to function.
———————————————
Elephants require Vitamin K for the liver to function.

People judge arguments like (4) to be stronger than arguments like (3), showing sensitivity to the greater diversity of the premise categories in (4), hippos and hamsters, despite the greater similarity between rhinos and elephants than between hamsters and elephants. Such a case would go against Sloutsky and Fisher's claim that similarity between premise and conclusion information is primary. Carey (1985) suggested that investigating diversity, like typicality, would be particularly informative about children's conceptual knowledge. Because Sloutsky and Fisher used only one premise for each argument, it was not possible to address this issue. The Osherson et al. study was conducted with adults, but it was followed with a study by Lopez, Gelman, Gutheil, and Smith (1992), who examined 9 phenomena with children of ages 5 to 9 (also see Gutheil & Gelman, 1997). Only 2 of these phenomena could be addressed with Sloutsky and Fisher's design, again because it did not involve multiple premises or varying the conclusion between arguments.

The second narrowing relates to people's ability to make inductive inferences about either individuals or categories. In Sloutsky and Fisher's (2004) studies, children were always asked about individuals, such as if one animal had a property would another animal have that property. In contrast, research with children and adults has generally examined inferences about categories, as in the above examples. It is unknown whether there are systematic differences between induction about individuals versus categories (Heit, 2000), but it seems possible that reasoning about specific individuals would be particularly influenced by similarity rather than by other variables (Blok, Newman, & Rips, 2005). On the other hand, it is possible that when children are asked about individuals, they will still use some knowledge about categories. Because the Sloutsky and Fisher studies asked only about individuals, it is unclear whether children used similarity and category knowledge differently in this study than in other studies of induction.

The third narrowing of scope concerns the use of pictures for presenting the materials and, more broadly, the exclusive focus on perceptual similarity. Here, the practices in previous research have been somewhat different for adults versus children. In virtually all studies of inductive reasoning in adults to date, materials have been presented verbally or in written form, as in the sentences in Arguments 3 and 4 above, rather than in pictures. Therefore, previous studies with adults have not examined perceptual similarity but rather similarity in more general terms. Although it is possible that an assessment of similarity between, say, hippos and elephants might include some memories of perceptual information, it is extremely likely that such an assessment would also include conceptual knowledge that is not derived from direct perceptual experience, for example, that both kinds of animals are warm-blooded and have brains. Previous studies with children have generally used a mixture of picture presentation and verbal (spoken) presentation in place of pictures. For example, Lopez et al. (1992) and Lo, Sides, Rozelle, and Osherson (2002) included some tasks in which pictures corresponding to premise categories were presented, and then children were asked about a general conclusion, whether "all animals" have some property of interest, without seeing a picture for "all animals." Lo et al. argued that presenting pictures of specific conclusion categories, namely, particular kinds of animals such as dogs or horses, might unduly encourage the use of similarity between premise and conclusion categories and hence would mask children's other tendencies. In general, previous developmental studies have shown that children do respond systematically when premise items are presented in pictures and a general conclusion category such as "all animals" is presented in spoken form, despite the impossibility of finding a perceptual overlap.

The account suggested by Sloutsky and Fisher (2004) might be preserved by assuming that when children hear a general conclusion like "all animals have *X*," they may remember a sample of exemplars from the relevant category, including perceptual information about each instance. Induction might then proceed via an assessment of the similarity between the premise items and the exemplars sampled from the conclusion category. As well as requiring a variety of additional assumptions (e.g., specifying a mechanism for memory retrieval and stipulating that only perceptual information is retrieved), there are problems with this argument. First, it is unclear how the SINC model would calculate similarity across multiple conclusion items. Second, like adults, children seem to understand that statements like "all animals have property *X*" do apply to all members of a category and not just to a subset of exemplars (Gelman, Star, & Flukes, 2002).

More generally, Sloutsky and Fisher's (2004) account of induction in terms of perceptual similarity seems to underestimate the complex role of similarity in cognitive processes and the difficulties in drawing a sharp distinction between perceptual and conceptual similarity (e.g., Goldstone & Son, 2005; Hahn & Chater, 1997; Murphy, 2004). Categorization can be based on perceptual features or nonperceptual features (e.g., Barsalou, 1991). Even the perceptual feature space over which similarity would be computed is not fixed but dependent on ontological knowledge (Booth & Waxman, 2002; Keil, 1995; but see Smith, Jones, Yoshida, & Colunga, 2003), task goals, and the feedback received during learning (Schyns, Goldstone, & Thibaut, 1998). For example, Keil (1995) reported that depending on whether an object is labeled as a *frog* or a *rock,* children use different perceptual features to extend the meaning of the word. By no means do we claim that stimuli based on words are inherently better than picture stimuli. However, when drawing conclusions about the role of perceptual similarity, in particular visual matching, the conclusions are going to be heavily dependent on whether all the stimuli are presented as pictures.

Finally, Sloutsky and Fisher (2004) provided an overly narrow view of children's induction through the exclusive use of living things and unfamiliar biological properties (e.g., "has a square heart"). Although this focus on biological categories and unfamiliar properties is common within the wider literature, it nevertheless restricts the kinds of questions that can be asked. Sloutsky and Fisher described the possibility that property knowledge would affect their results as "remote" (p. 185). Yet notably, studies that have varied properties and category domains have found that even preschool children can make inferences on the basis of their knowledge about ontological kinds or about the specific property in question (see Gelman, 2003, for a review). In the case of biological kinds, preschool children do not generalize arbitrary properties, for example, "fell on the floor this morning," from premise to conclusion categories even when there is a strong perceptual similarity between them (Gelman, 1988). They also do not draw inferences about shared properties when two animals are labeled with a term that describes a transient state, for example, "this is sleepy" (Gelman & Coley, 1990). In addition, there is evidence that children distinguish between compositional and functional information when drawing inferences about different properties (Kalish & Gelman, 1992; Nguyen & Murphy, 2003). In short, children's knowledge about properties can overcome effects of the perceptual overlap between premise and conclusion stimuli (see Heit & Rubinstein, 1994, for similar findings with adults). This point implies a role for conceptual knowledge in children's induction that goes well beyond Sloutsky and Fisher's (2004) account based on a limited form of perceptual similarity.

In summary, there is a rich set of phenomena in induction by children and adults. Both children and adults clearly can perform induction on information other than perceptual similarity, for example, when pictures are not presented. Sloutsky and Fisher (2004) used a triad task involving pictures of specific individuals and unfamiliar biological properties, and as such examined a narrow range of phenomena for which perceptual similarity might be used, but not in a way that is representative of the broader context of induction. From these studies, it is not possible to draw conclusions about the role of perceptual similarity in the majority of inductive reasoning phenomena. In many cases involving induction without pictures or induction about general conclusion categories, it simply would not be possible to use perceptual similarity. Likewise, it may be possible to explain Sloutsky and Fisher's results without further assumptions about category structure or taxonomic knowledge, but the simple nature of the experimental design precludes the investigation of many phenomena that could provide more definitive evidence regarding children's sensitivity to taxonomic knowledge and other conceptual information.

Models of Inductive Reasoning

We now turn from the wider picture of experiments to the wider picture of models of induction. The SINC model proposes systematic relations between similarity and judgments of inductive strength. The first point to make about the SINC model is that it was presented only as an account of tasks with a triad design, in which two single-premise arguments are compared. As such, the SINC model differs from previous models of inductive reasoning (Heit, 1998; Osherson et al., 1990; Rips, 1975; Sloman, 1993), which have been applied to a much wider variety of phenomena, such as reasoning based on more than one premise. In this section, we compare the SINC model to other models' accounts of the triad task. The main question addressed is whether the results accounted for by the new similarity-based model, SINC, can also be accounted for by previous models, similarity based or not.

The SINC model uses a multiplicative rule for calculating similarity (Medin & Schaffer, 1978), as shown in Equation 1:

$$\mathrm{Sim}(i, j) = S^{N-k}. \tag{1}$$

Here, the similarity (Sim) between pictures i and j depends on the number of features compared, N, the number of matching features, k, and a free parameter, S, taking on a value between 0 and 1. In general, as there are fewer matches or more mismatches, the similarity rule gives a lower value. This rule is adapted when category labels are presented along with the pictures, as in Equation 2. In effect, the category label is treated as another feature (for similar assumptions, see Estes, 1994; Heit, 1992; Love, Medin, & Gureckis, 2004):

$$\mathrm{Sim}(i, j) = W^{1-L}S^{N-k}. \tag{2}$$

Here, W is another free parameter between 0 and 1, and L is given a value of 1 when the two pictures have matching labels and a value of 0 when there is a label mismatch. Hence, label mismatches will tend to reduce similarity.

The SINC model adopts Luce's (1963) choice rule to predict a response based on similarity. As the similarity between the target and Test B increases relative to the similarity between the target and Test A, so does the probability of choosing Test B in the induction task. Equation 3 shows the probability of choosing B over A, given the target item T:

$$P(\mathrm{B}) = \frac{\mathrm{Sim}(\mathrm{T, B})}{\mathrm{Sim}(\mathrm{T, B}) + \mathrm{Sim}(\mathrm{T, A})}. \quad (3)$$

A representative application of the SINC model is shown in Figure 1. The solid lines show the proportion of Test B choices for various stimuli for induction and categorization judgments. The *x*-axis shows the relative similarity between the target and Test A versus Test B, in effect Sim(T, A)/Sim(T, B). The similarity ratios were estimated from a separate norming experiment that used unlabeled pictures. In general, the proportion of B choices falls as the similarity ratio increases, that is, as the target is increasingly more similar to Test A rather than Test B. This figure shows the results for conditions in which the child is or is not told the category labels. In the label condition, in which the target and Test B share a category label, the proportion of Test B choices is elevated. Finally, the dashed lines in Figure 1 show the predictions of the SINC model, with reasonable assumptions about the values of the *S* and *W* parameters (.5 and .35, respectively). Note that the SINC model does not distinguish between induction judgments and categorization judgments: They are modeled the same way.

Rips (1975) Model

Now that the SINC model has been presented, we consider previous models of induction. Our aim is to present variants of previous models that are transparently related to SINC and account for the same set of results with the triad design. It is likely that previous models could also be applied in other ways. The first model considered is perhaps the original model relating similarity and inductive strength. The model by Rips (1975) was applied to adults' reasoning about verbal statements, such as "If all the ducks on an island have a certain disease, what proportion of the eagles would have this disease?" According to this model, inductive strength depends on the typicality of the premise item (e.g., whether ducks are close to the prototypical bird) and the similarity between the premise and conclusion items (ducks and eagles). The Rips model assumes a critical role for typicality, defined in terms of the distance between any item and the superordinate category's central tendency. In contrast, the SINC model does not consider typicality. The Sloutsky and Fisher (2004) stimuli possibly varied in typicality, but results were averaged over items, so it is not possible to assess typicality effects. Therefore, for application to these studies, we focus on the similarity component of the Rips model.

In Rips (1975), similarity was derived from distance in a multidimensional scaling solution, based on adults' similarity judgments for pairs of animals. The Rips model does not distinguish among similarity, perceptual similarity, and similarity based on category labels. The original stimuli were presented as words rather than pictures, so it is extremely likely that there was some influence of nonperceptual information. Strictly speaking, the Rips model was used only for predicting inductive strength judgments from distance measures: It was assumed that inductive strength was a linear decreasing function of distance between the target and test items. The Rips model did not use similarity directly, but there are standard established relations between distance and similarity (e.g., Nosofsky, 1988), as in Equation 4:

$$\mathrm{Sim}(i, j) = e^{-cD(i,j)}. \quad (4)$$

In this equation, $D(i, j)$ is the distance between items i and j in multidimensional space, and c is a free parameter with a positive value. The Rips model does not make any claims about the influence of shared category labels on judged similarity. However, the Rips model could be easily adapted to make the same assump-

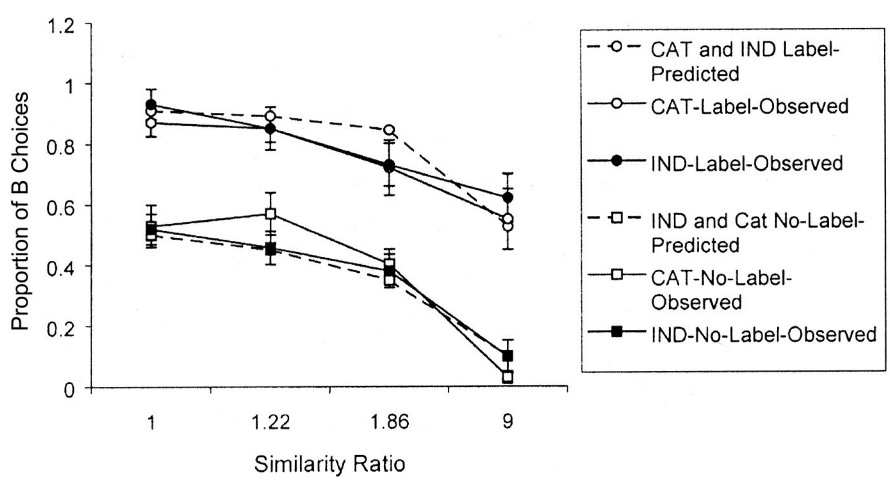

Figure 1. Predicted and observed probabilities of B choices as a function of similarity ratio and labeling in induction and categorization tasks for Experiment 2. IND = induction; CAT = categorization. From "Induction and Categorization in Young Children: A Similarity-Based Model," by V. M. Sloutsky and A. V. Fisher, 2004, *Journal of Experimental Psychology: General, 133,* p. 175. Copyright 2004 by the American Psychological Association.

tion as the SINC model, namely, that shared category labels also contribute to similarity. When category labels are present, Equation 5 could be applied, with W and L having the same roles as in Equation 2:

$$\mathrm{Sim}(i, j) = W^{1-L}e^{-cD(i,j)}. \quad (5)$$

Finally, Equation 3 could be used to make forced choice predictions in a triad task. In sum, there is an equivalence between the Rips (1975) model and the SINC model. Any pattern of results in the triad task that can be accounted for by the SINC model can be accounted for by the Rips model, and vice versa. This equivalence depends on some assumptions. One is the standard relation between psychological distance and similarity. Another is that the Rips model could make the same assumption as SINC regarding the influence of category labels on assessment of similarity.

Osherson et al. (1990) Model

The most influential and widely cited model of induction is Osherson et al.'s (1990) similarity-coverage model. This model was applied to 11 phenomena and, as such, has a wider scope of application than does the SINC model. As we show below, for a simple triad design in which only similarity is manipulated, the SINC model and the Osherson et al. model make the same predictions. The first point to state about the Osherson et al. model is that it predicts inductive strength from similarity. There is no assumption that similarity corresponds to perceptual similarity. When this model was applied to reasoning by adults, the experimental materials were presented as words rather than pictures, and it is doubtless that nonperceptual information had an influence. When this model was applied to reasoning by children (Lopez et al., 1992), the experiments used pictures as well as presentations by verbal description alone (e.g., "all animals"), in which case children had to use nonvisual information. The Osherson et al. model does not have an equation for calculating similarity on the basis of feature matches, so instead subjects' responses in a similarity rating task are used directly.

The second point to make about this model is that, like the Rips (1975) model, it has two components, corresponding to similarity and coverage. These components are somewhat complex for arguments with multiple premises, but they are straightforward for single-premise arguments. For premise i and conclusion j, the similarity is simply $S(i, j)$, taken from adults' judgments of similarity. The coverage component depends on generating a superordinate category including all items mentioned in the inductive argument. In Sloutsky and Fisher's (2004) experiments, the superordinate category could be mammals, for example, when a cat and a dog were presented; animals, for example, when a bird and a bat were presented; or living things, for example, when a bug and a leaf were presented. The coverage of an argument is simple to calculate for single-premise arguments: It is just the similarity of the premise item to members of the superordinate, in effect the item's typicality.

Hence, for an inductive argument with a single-premise item, the Osherson et al. (1990) model is shown in Equation 6:

$$\mathrm{Strength}(i, j) = \alpha S(i, j) + (1 - \alpha)S\{i, \mathrm{Superordinate}(i, j)\}. \quad (6)$$

Here, the inductive strength of an argument concerning premise i and conclusion j depends on a weighted sum of the judged similarity between i and j, $S(i, j)$, and the typicality of i, which is simply the judged similarity of i to members of its superordinate category. As noted in the discussion of the Rips (1975) model, Sloutsky and Fisher's (2004) design did not assess the role of typicality, and any typicality effect would be averaged across items, in effect just adding a constant to the strength of each argument. Hence, for application here, the similarity component of the Osherson et al. model is again the focus. Similarity in the Osherson et al. model can easily be converted to the currency of the SINC model, as shown in Equations 7 and 8. Equation 7 is used for tasks in which category labels are not given, and Equation 8 is used when category labels are given.

$$\mathrm{Sim}(i, j) = S(i, j) \quad (7)$$

$$\mathrm{Sim}(i, j) = W^{1-L}S(i, j) \quad (8)$$

Finally, the response rule in Equation 3 would be used.

In sum, any results in the triad task that can be accounted for by the SINC model can be accounted for by the Osherson et al. (1990) model, and vice versa. This equivalence depends on the assumption that shared category labels would affect similarity in the same way as in the SINC model.

Sloman (1993) Model

Sloman's (1993) feature-based induction model is an alternative to the Osherson et al. (1990) model. Both models account for a wide range of phenomena in inductive reasoning. Perhaps the most important difference between the two models is that the Sloman model does not have a component that explicitly calculates coverage with respect to a superordinate category. Although Sloutsky and Fisher (2004) noted that this difference with regard to use of taxonomic knowledge is important, on the other hand the coverage component plays no role in the triad task. When comparing two single-premise arguments with a common conclusion item, the differences among the Sloman model, the Osherson et al. model, and the SINC model should be minimal.

In the Sloman (1993) model, inductive strength depends on similarity in terms of feature overlap. These features are not limited to perceptual features. In Sloman's application of the model, the features were derived from a verbal feature rating task. This model has a procedure for encoding multiple-premise arguments in terms of training a feedforward connectionist network. However, the model can be described in a much simpler way for single-premise arguments, as shown in Equation 9:

$$A(i, j) = \frac{\mathbf{F}(i) \cdot \mathbf{F}(j)}{|\mathbf{F}(j)|^2}. \quad (9)$$

In the numerator of this equation for A, or activation, the dot product function is a measure of the extent to which the premise and conclusion items have overlapping features. $\mathbf{F}(i)$ and $\mathbf{F}(j)$ refer to vectors of features for the premise and conclusion items, respectively. Each vector consists of an ordered set of feature values between 0 and 1, corresponding to degree of presence of this feature for the item. The dot product just takes the sum of the products of pairs of values from the two vectors, so that the sum will tend to increase only when both items have a nonzero value for that feature. The denominator of this equation refers to the magnitude of the conclusion vector. In effect, the magnitude just measures the number of known features for the conclusion item. For example, more typical items might have more known features (see also Hampton & Cannon, 2004). However, in the triad task, the same conclusion item was used for two arguments, so the magnitude of the conclusion item would have no effect on the relative strengths of these two arguments.

Given these assumptions, judged similarity in the Sloman (1993) model can be converted to the currency of the SINC model, as shown in Equations 10 and 11. Equation 10 is used for tasks in which category labels are not given, and Equation 11 is used when category labels are given.

$$\mathrm{Sim}(i, j) = A(i, j) \tag{10}$$

$$\mathrm{Sim}(i, j) = W^{1-L}A(i, j) \tag{11}$$

Finally, the response rule in Equation 3 would be used. The conclusion then is the same as for the Rips (1975) and Osherson et al. (1990) models: Any results in the triad task that can be accounted for by the SINC model can be accounted for by the Sloman (1993) model, and vice versa.

Heit (1998) Model

Unlike the other previous models, the Heit (1998) model of inductive reasoning is based not on similarity but on hypothesis spaces and Bayesian statistics (see also Tenenbaum & Griffiths, 2001). This model can address a wide range of phenomena, including putting together information from multiple premises; however, the model itself takes a simple form for single-premise arguments, as in the following example:

$$(5)\ \frac{\text{Cows have Property P}}{\text{Horses have Property P}}$$

When Argument 5 is evaluated, the goal is to estimate the range of property *P*. Which animals have property *P* and which do not? To reach this goal, people would initially assume that property *P* is distributed like other known properties of animals. There are four possibilities in this hypothesis space. Property *P* might be a property that is true of both cows and horses, like "has a heart." Or it might be a property that is true of cows but not horses, such as "makes a mooing sound." Or it could be a property that is true of horses but not cows, such as "commonly used for racing." Finally, it could be a property that is true of neither cows nor horses, such as "lives under the sea." The initial degrees of belief in these four hypotheses can be labeled *E, F, G,* and *H,* respectively.

After this initial assessment, the premise information is taken into account, namely, that cows do have property *P.* At this point, only two hypotheses are still viable. Namely, property *P* could be a cow-and-horse property, or it could be a cow-and-not-horse property. The strength of the conclusion that horses have this property is just the relative degrees of beliefs in these two hypotheses, as shown in Equation 12:

$$S(i, j) = \frac{E}{E + F}. \tag{12}$$

In other words, this argument will seem strong to the extent that there are many known properties shared by cows and horses, represented by the value *E,* and will seem weak to the extent that there are many known properties of cows not shared by horses, represented by *F.*

By taking account of beliefs about known properties, this model can predict similarity effects. For example, consider the following argument:

$$(6)\ \frac{\text{Mice have Property P}}{\text{Horses have Property P}}$$

For Argument 6, the knowledge of shared properties between mice and horses, represented by *E,* will be lower than that for Argument 5. Knowledge of properties of mice but not horses, represented by *F,* would be greater for Argument 6 than for Argument 5. Hence, from Equation 12, the argument concerning dissimilar items, mice and horses, would be weaker.

Indeed, the strength function in Equation 12 can be thought of as a measure of similarity between the premise and conclusion items. Hence, for application to Sloutsky and Fisher's (2004) experiments, inductive strength in the Heit (1998) model can be converted to the currency of the SINC model. Equation 13 is used for experimental conditions in which category labels are not given, and Equation 14 is used for conditions with category labels.

$$\mathrm{Sim}(i, j) = S(i, j) \tag{13}$$

$$\mathrm{Sim}(i, j) = W^{1-L}S(i, j) \tag{14}$$

Finally, the response rule in Equation 3 would be used. In sum, any results in the triad task that can be accounted for by the SINC model can be accounted for by the Heit (1998) model, by estimating *E* and *F* values corresponding to degrees of belief about feature overlap.

Summary

Although there are important differences among the four previous models, these differences tend to vanish for simple triad designs. These models have additional mechanisms for assessing typicality, computing coverage, or putting together information from multiple premises, but these mechanisms were not addressed by Sloutsky and Fisher's (2004) studies. Although the SINC model can account for the results of these studies, so can the previous models. One important contribution made by SINC is the assumption that category label matches or mismatches can modify similarity. However, this assumption could be added to any previous model of induction as well. To experimentally distinguish SINC from previous models, it would be necessary for one to use more complex designs, such as looking at how people use multiple premises. However, the SINC model is at present only formulated for single-premise arguments and hence cannot account for any of the important multiple-premise phenomena addressed by other models of induction.

Another contribution made by the SINC model is the idea of predicting induction and categorization judgments from the same set of equations. Previous models had been applied only to induction judgments, so these comparisons among models suggest that previous models of induction might also be applied to categorization. Although Sloutsky and Fisher (2004) found only a correspondence between induction and categorization for a simple triad design with children, it is possible that the correspondence extends more generally. For example, Rehder and Hastie (2001, 2004) found fairly close correspondences between induction and categorization judgments over a wider range of stimulus designs for adult subjects. It remains to be seen how far the parallels between induction and categorization can be taken. Just focusing on the Osherson et al. (1990) model, recall that the full version of this model was applied to 11 different induction phenomena. Are there 11 corresponding phenomena in categorization, and could the same model account for the details of the results?

Relations Between Induction and Recognition

In general, Sloutsky and Fisher (2004) examined induction and categorization. However, in Experiment 5, the relation between induction and recognition was assessed, by having subjects make induction judgments followed by recognition judgments. For example, children and adults were shown a set of animal pictures, illustrating that a set of cats has some property but a set of bears does not have that property. Both children and adults were successful on this induction task. Then, there was an unexpected recognition memory test. Children could distinguish pictures of cats that had been previously presented from new pictures of cats, but adult recognition memory for this distinction was poor, as if the adults had only represented the categories. The adult results are particularly surprising in light of adults' usually impressive recognition memory for pictures (Shepard, 1967). Sloutsky and Fisher interpreted the results in terms of children performing induction in a similarity-based manner, in comparison to adults performing induction in a more category-based manner.

What are the implications of these results for models of induction? The first point to make is that none of the models of induction, including SINC, have been implemented as models of recognition. Therefore, any predictions made for these models are necessarily speculative. The Sloman (1993) model is explicitly not a category-based model, so any results indicating that children's judgments are similarity based might be just as compatible with the Sloman model as with SINC. The Osherson et al. (1990) model is supposed to be category based, and as part of its coverage mechanism the model generates a superordinate category. It is tempting to think that generating a superordinate would lead to the false recognitions shown by adult subjects, such as confusing one cat with another. However, examination of the Osherson et al. model does not lead to an easy explanation. The superordinate generated by this model would have to include all of the stimulus items, such as cats, bears, and birds. Hence, the superordinate would be "animals." If each picture of a cat, bear, or bird were simply encoded as an "animal," this would lead to the absurd prediction that adults would also falsely recognize other animals such as dogs and fish. Hence, the superordinate category generated by the Osherson et al. model should not be taken as the basis for predicting recognition judgments.

Although the SINC model has not been implemented as a model of recognition, it might be modified to make predictions for recognition tasks. Estes (1994) noted that similarity-based categorization models can also be used as models of recognition. In effect, a recognition judgment is choosing whether a stimulus belongs to the category of old, previously observed items or the category of new items. On this basis, Equation 3 can be recast as Equation 15:

$$P(\text{say } x \text{ is "old"}) = \frac{\text{Sim}(x, \text{old items})}{\text{Sim}(x, \text{old items}) + \text{Sim}(x, \text{new items})}. \quad (15)$$

In this equation, Sim(*x,* old items) is the total similarity (cf. Jones & Heit, 1993) between stimulus *x* and items retrieved from memory. Because the members of the new category are usually not known, Sim(*x,* new items) is normally treated as a parameter of the model, which might vary between experimental conditions.

This equation has the useful property of expressing a relation between similarity and recognition. For example, if a person observes a picture of a house, the person might falsely recognize a picture of another house with a similar appearance. Furthermore, the same variables should affect both similarity and recognition. If the person is induced to pay great attention to category labels and to ignore perceptual similarity, the person would tend to falsely recognize other pictures of houses, even if they look different from the studied item.

Could the SINC model, modified in this manner to address recognition, account for the results of Sloutsky and Fisher (2004)? For example, could the different similarity judgments (different Sim functions) for children and adults be used to predict different patterns of recognition? Experiment 1 was a similarity judgment task comparing children and adults, and the main result was that children were influenced by shared category labels but adults were not. In contrast, Experiment 5 had the opposite pattern of results. Children's recognition judgments were unaffected by category membership, but adults showed a distinctly categorical pattern. Clearly, for adults, similarity judgments could not be used to directly predict recognition. For children, it is notable that category labels were presented in spoken form for Experiment 1, but in Experiment 5, the stimuli were pretested to be easily labeled by children, although the labels were not spoken by the experimenter. It is unclear what effect self-generated category labels would have on these children's judgments. The main point though is that the similarity results of Experiment 1 could not be used to predict the whole pattern of recognition results in Experiment 5, using an extension of SINC based on Equation 15.

In sum, Experiment 5 is an important contribution because it begins to show how relations between induction and recognition could be studied. However, it is not yet known how SINC or other models of induction would address the results, because they have not been implemented as recognition models. The most straightforward extension of SINC, along the lines of Estes (1994), would have obvious difficulties, such as in simultaneously accounting for Experiments 1 and 5. It is no doubt too early to develop a model of induction and recognition based on these experiments. It would be important to assess a wider range of phenomena in induction and examine their effects on recognition memory performance before trying to create a model encompassing both abilities (Hayes & Heit, 2004).

Conclusions

Perhaps the most important conclusion is that the Sloutsky and Fisher (2004) article is ambitious, yet more ambition is needed. It is important to look for connections among categorization, induction, recognition, and similarity. Devising experiments in which two or more of these issues are addressed, conducting these experiments with children and adults, and developing mathematical models to account for the results seem like an extremely valuable strategy. Furthermore, some of the key assumptions made by Sloutsky and Fisher, such as that shared category labels could contribute to judged similarity and that the same model could be used for both induction and categorization, are assumptions that could be usefully adopted by other researchers in this area.

However, the Sloutsky and Fisher (2004) studies also had some serious limitations. Because of the focus on a narrow range of phenomena, it is impossible to distinguish the SINC model from other models of induction. A wider range of experimental designs would be needed, because other models can predict the same results as SINC for these experiments. The main purpose of Sloutsky and Fisher's studies was to assess whether children perform induction in a category-based manner, yet the simple triad design does not allow the examination of phenomena that would be most revealing about categorical knowledge. The other limitation was the focus on perceptual similarity, which seems heavily linked to the exclusive use of picture stimuli. Yet models of adult induction routinely assume the use of conceptual information for the assessment of similarity. Even young children can make inductive judgments in regard to a premise item shown as a picture and a conclusion item shown without a picture, such as "all animals," in which case it would be impossible to assess the direct visual overlap between the premise and conclusion items. Regardless of how Sloutsky and Fisher's studies are interpreted, they do not show that perceptual similarity on its own is a viable account of the complete range of past results in children's induction.

Therefore, we hope that future research will build on the contributions of Sloutsky and Fisher (2004) regarding relations among categorization, induction, recognition, and similarity but at the same time go beyond this work's boundaries. One particularly exciting possibility is that a common model could be developed addressing a range of corresponding phenomena in categorization, induction, and recognition in children and adults.

References

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

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 1–64). San Diego, CA: Academic Press.

Blok, S., Newman, G., & Rips, L. J. (2005). Individuals and their concepts. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 127–149). Washington, DC: American Psychological Association.

Booth, A. E., & Waxman, S. R. (2002). Word learning is 'smart': Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition, 84,* B11–B22.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: Bradford Books.

Estes, W. K. (1994). *Classification and cognition.* New York: Oxford University Press.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20,* 65–95.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford, England: Oxford University Press.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26,* 796–804.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23,* 183–209.

Gelman, S. A., Star, J., & Flukes, J. (2002). Children's use of generics in inductive inferences. *Journal of Cognition and Development, 3,* 179–199.

Goldstone, R. L., & Son, J. Y. (2005). Similarity. In K. J. Holyoak & R. C. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 13–36). Cambridge, England: Cambridge University Press.

Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64,* 159–174.

Hahn, U., & Chater, N. (1997). Concepts and similarity. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 43–92). London: Psychology Press.

Hampton, J. A., & Cannon, I. (2004). Category-based induction: An effect of conclusion typicality. *Memory & Cognition, 32,* 235–243.

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

Hayes, B. K., & Heit, E. (2004). Why learning and development can lead to poorer recognition memory. *Trends in Cognitive Sciences, 8,* 337–339.

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

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7,* 569–592.

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

Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology, 43,* 243–273.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 411–422.

Jones, C. M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 799–812.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classifications and children's category-based inductions. *Child Development, 63,* 1536–1557.

Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.). *Causal cognition: A multidisciplinary debate* (pp. 234–267). Oxford, England: Oxford University Press.

Krascum, R. M., & Andrews, S. (1998). The effects of theories on children's acquisition of family resemblance categories. *Child Development, 69,* 333–346.

Lamberts, K., Brockdorff, N., &. Heit, E. (2003). Feature-sampling and random-walk models of individual-stimulus recognition. *Journal of Experimental Psychology: General, 132,* 351–378.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science, 16,* 181–206.

Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development, 63,* 1070–1090.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309–332.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–187). New York: Wiley.

Markman, E. M. (1989). *Categorization and naming in children.* Cambridge, MA: MIT Press.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10,* 517–532.

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

Murphy, G. L. (2004). On the conceptual–perceptual divide in early concepts. *Developmental Science, 7,* 513–515.

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

Nguyen, S. P., & Murphy, G. L. (2003). An apple is more than just a fruit: Cross-classification in children's concepts. *Child Development, 74,* 1783–1806.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700–708.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130,* 323–360.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition, 91,* 113–153.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665–681.

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

Schyns, P. G., Goldstone, R. L., & Thibaut, J. (1998). Development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1–54.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231–280.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133,* 166–188.

Smith, L. B., Jones, S., Yoshida, H., & Colunga, E. (2003). Whose DAM account? Attentional learning can explain Booth and Waxman. *Cognition, 87,* 209–213.

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

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.

Received August 11, 2004
Revision received December 15, 2004
Accepted December 28, 2004 ■