Journal of Experimental Psychology: General
2000, Vol. 129, No. 2, 220–241
Copyright 2000 by the American Psychological Association, Inc.
0096-3445/00/$5.00 DOI: 10.1037//0096-3445.129.2.220

Why Do Categories Affect Stimulus Judgment?

Janellen Huttenlocher, Larry V. Hedges, and Jack L. Vevea
University of Chicago

The authors tested a model of category effects on stimulus judgment. The model holds that the goal of stimulus judgment is to achieve high accuracy. For this reason, people place inexactly represented stimuli in the context of prior information, captured in categories, combining inexact fine-grain stimulus values with prior (category) information. This process can be likened to a Bayesian statistical procedure designed to maximize the average accuracy of estimation. If people follow the proposed procedure to maximize accuracy, their estimates should be affected by the distribution of instances in a category. In the present experiments, participants reproduced one-dimensional stimuli. Different prior distributions were presented. The experiments verified that people's stimulus estimates are affected by variations in a prior distribution in such a manner as to increase the accuracy of their stimulus reproductions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

It is well known that categorization affects stimulus judgment. Two stimuli that are actually equidistant may be judged as more similar or may be harder to discriminate if they are from a common category than if they are from different categories, and an individual stimulus from a particular category may be judged closer to the center of that category than it actually is. There is an extensive literature starting early in the century, and continuing to the present, showing bias due to categories in many domains. For example, the social categories that people use may affect their judgment of individuals, the set (category) of experimental stimuli presented to people may affect their judgment of particular stimuli, and so forth (e.g., Bartlett, 1932; Brewer & Nakamura, 1984; Hollingworth, 1910; Poulton, 1989; Tajfel, 1959).

In the present article, we test a model that holds that category effects on stimulus judgments arise in pursuit of an adaptive goal—to maximize accuracy. Anderson (1990, 1991) has suggested that establishing what goals an organism is pursuing should be the first step in cognitive modeling. Models of goals should precede detailed models of cognitive processing, he argues, because goals constrain processes. We propose a Bayesian model of stimulus judgment in which people use prior information, in the form of categories, to adjust inexactly represented stimuli; this gives rise to bias but improves accuracy by reducing the variability of estimates. Indeed, if the goal of stimulus judgment is to maximize average accuracy, it would be irrational for the cognitive system *not* to produce biased estimates.

Recently, a variety of findings, originally interpreted as showing bias and flawed reasoning, have been reexamined to determine whether people might actually be using rational processes, appropriate to their goals. Oaksford and Chater (1996) have reinterpreted apparent errors on deductive reasoning tasks (e.g., Wason 1968), showing that people actually may treat the tasks as inductive problems where "a Bayesian approach to optimal data selection" is used to maximize "expected information gain" (p. 382). Earlier, Kahneman and Tversky (1972) investigated whether people use Bayesian procedures to improve the accuracy of their judgments under uncertainty. In categorization tasks, where it is uncertain what category a stimulus belongs to, differences in the likelihood of the alternatives should be taken account of in decision making. Kahneman and Tversky found that the relative frequencies of alternatives were often ignored (*base rate neglect*), seemingly indicating that humans are not rational operators. However, in an extensive review of the literature, Koehler (1996) found that base rate neglect is not the rule; in many tasks, people do use base rates in making decisions. Kruschke (1996) presented evidence that even in cases of apparent base rate neglect, it can be shown that people actually are sensitive to base rates in category decisions. Nosofsky (1998) has shown that his exemplar model of how people decide what category a stimulus belongs to actually constitutes an explicit Bayesian model, even though the model arose from other considerations.

Although the use of relative frequencies *across* categories to improve the accuracy of category membership judgments has been explored, the possibility that relative frequencies *within* a category are used in a parallel fashion in judgments of individual stimulus values has not been systematically

Editor's Note: Eleanor M. Saffran served as the action editor for this article.—NSN

Janellen Huttenlocher, Larry V. Hedges, and Jack L. Vevea, Department of Psychology, University of Chicago. Jack L. Vevea is now at the Department of Psychology, University of North Carolina.

We thank Judith Avrahami, Lawrence Barsalou, Beth Crawford, and Yakov Kareev for their helpful comments on drafts of this article and Peder Hans Engebretson for his help in early work on this project. Special thanks are due to William Goldstein for illuminating discussions of the psychophysical literature. This research was supported in part by Grant R01 MH45402 from the National Institute of Mental Health.

Correspondence concerning this article should be addressed to Janellen Huttenlocher, Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, Illinois 60637.

explored. Yet it is widely accepted that, at least in certain ways, prior information about relative frequency within categories *is* used adaptively; notably, correlational structure is used to infer the unobservable features of stimuli (Rosch, 1975; e.g., the more green a liquid is, the more likely it is to be poisonous). In general, relative frequency is potentially functional information. Consistent with this view, Murphy and Ross (1994) found that people even preserve relative frequency for nondefining attributes of categories. Here we examine whether within-category relative frequencies are used to increase accuracy in the judgment of inexactly represented stimuli from those categories.

Our model is a precisely specified Bayesian model. It holds that in pursuing the goal of maximizing accuracy, people use prior information in estimating stimulus values that are represented inexactly. Prior information is incorporated into decision making in the form of an explicit prior distribution, and the inexactness of the fine-grained information is incorporated as a sampling distribution. Given a category (an explicit prior distribution) and an inexact stimulus value (a sampling distribution describing the uncertainty of current data), Bayes's theorem provides a method for combining the information to provide estimates with certain optimal properties. A posterior distribution summarizes uncertainty after combining the uncertain data and the prior information. The mean of this posterior distribution is called the Bayesian estimate; it has the property of being the "most accurate" estimate in the sense that it minimizes average error.

In our model, category effects arise at estimation when people combine fine-grain and category information to achieve high accuracy of judgment. Other recent approaches posit that category effects arise at encoding. Category learning, it is argued, alters the way stimulus dimensions are represented in memory. For example, Nosofsky (1986) proposed that category learning leads to the "stretching" of a category's dimensions relative to other dimensions. Goldstone (1994) proposed that, in addition, category learning stretches selected regions of a single dimension when that dimension is divided into more than one category. The notion of stretching at category boundaries has its roots in earlier work explaining why phonemes are easier to discriminate when they are from different categories than from the same category, a phenomenon described as *categorical perception* (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Goldstone obtained parallel effects on discrimination after teaching people categories, and he too described the effect as occurring at perception or encoding. If category learning alters encoding because it changes the psychological distances among physically equidistant stimuli, estimates of stimuli also should be affected by such learning. The question of whether observed biases in estimation can be explained in terms of effects of stretching at boundaries on encoding is discussed at the end of the article.

The Category Adjustment Model

This article concerns inductive categories—that is, categories formed from observed sets of stimuli. The model deals with statistical aspects of the representation of stimuli and categories. We treat a stimulus as a point on a set of dimensions, and a category as a distribution of stimuli. Although inductive categories are treated as statistical distributions in our model, we recognize that this is an incomplete characterization of such categories. Inductive categories tend to capture statistical regularities that matter for human purposes, and they are often embedded in general theories (e.g., Goodman, 1972; Murphy & Medin, 1985). Hence in modeling categories as distributions of observed sets of stimuli, one should remember that some categories also incorporate expectations (general principles, theories) about what the new distribution should be like (e.g., where boundaries are located, distribution shape, etc.). For example, a person might set an upper boundary for the width of trucks as some proportion of the presumed width of highway lanes.[1]

In this article we have chosen categories that, it seems, would incorporate only distributional information, allowing us to evaluate a critical prediction of our Bayesian model: If people use prior distributions to increase the accuracy of their judgments of stimuli, then the characteristics of those distributions should affect estimation in the ways we describe in this article. We test the predictions using a simple task. We present, one at a time, unidimensional stimuli that vary over a bounded range. After each stimulus is removed, people reproduce it, providing us with information about the bias and variability of their estimates. By varying the distribution in which exactly the same stimuli are embedded, we can examine whether estimates of those stimuli vary as they should if people follow procedures to maximize the average accuracy of estimates.

We first provide the background for the predictions tested in this article by summarizing our earlier work, laying out the model's assumptions about how fine-grain stimuli and categories are represented and how information at these two levels is combined in processing (Huttenlocher & Hedges, 1994; Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Prohaska, 1988). (Note that the assumptions about representation and processing are skeletal, serving only to characterize the logic of people's strategy sufficiently to allow us to test our model of the hypothesized goals of stimulus judgment.) Then we test an important set of predictions of the model; these concern the ways variations in the nature of a prior distribution should affect stimulus estimation. We test these predictions in three studies.

Assumptions of the Model

Representation

Fine-grain stimulus values. A stimulus in the model consists of a value on a set of dimensions—for inanimate

[1] In fact, some categories are not induced from sets of instances at all. For example, spatial categories may be formed by treating the axes of symmetry of a figure as boundaries (Huttenlocher et al., 1991), or temporal categories may be formed by treating the start and end dates of academic terms as boundaries (Huttenlocher et al., 1988).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

objects, their physical characteristics (height, etc.), or for animates, their personal characteristics (intelligence, etc.)—but the mental representation of a stimulus generally is imprecise. Hence we treat stimulus representation as an area of inexactness around a stimulus value. In cases where people's judgments of stimuli fail to correspond to physically measured values (i.e., where judgment is biased), many investigators represent the stimuli using nonveridical psychological scales. In our model, we instead derive nonveridicality of stimulus judgment from physically measured stimulus values, explaining the nonveridicality associated with categories by positing that people combine representations at two levels of detail. We also use physically measured values to express the nonveridicality associated with dimensions that increase in magnitude (e.g., length and loudness). For such dimensions, the discriminability of pairs that are equidistant in physical units decreases as magnitude increases. The traditional use of a psychological scale (log scale) equalizes discriminability; however, we use an equivalent representation that preserves physically measured values, while recognizing that variability increases with stimulus magnitude (Huttenlocher, Hedges, & Bradburn, 1990).

Categories. A category is treated as a structure consisting of a bounded range of values along a set of relevant stimulus dimensions. The classic notion of categories is that boundaries are exact and that all stimuli within the boundaries are equivalently good members. However, this clearly is not true for categories induced from a set of stimuli. Such categories capture the distribution of stimuli in a set and can be described in terms of summary statistics (Anderson, 1991; Ashby & Lee, 1991; Fried & Holyoak, 1984; Homa, 1984). Such inductive categories have been described as having a "graded" structure, with instances that vary from good (near a central value, which is sometimes described as a "prototype") to poor, and boundaries that are uncertain ("fuzzy"; e.g., Kay & McDaniel, 1978; Rosch, 1975). Although the boundaries of inductive categories are uncertain, they fill an essential function of categories; that is, they are "projectible" (Quinton, 1957), supporting decisions as to whether a new stimulus is a member.

The categories used in experimental studies are like transient naturally occurring categories that capture regularities in particular contexts. For example, a person who is assembling a desk may form a transient category of the sizes of the screws in the kit. Such categories do not consist of sets of unrelated items that serve temporary goals like ad hoc categories such as "things to take out of the house in a fire" (Barsalou, 1983). Rather, they are based on stimulus characteristics that allow decisions about the membership of new stimuli (despite uncertainty near boundaries).

In category learning experiments, a paradigmatic design is to present two adjoining categories, X and Y; the test of acquisition is the ability to identify those stimuli as Xs or Ys, that is, to make a category judgment. The learning task involves establishing what differentiates Xs from Ys. For this task, models have varied in their assumptions about what is learned. Some models posit deterministic boundaries between categories, with errors arising because of inexactness of stimulus perception (Ashby & Lee, 1992; Maddox & Ashby, 1993). Other models posit that categorization of a new stimulus depends on an assessment of similarity to members of one category or another. In this formulation, the boundary between categories must be treated as uncertain (Kalish & Kruschke, 1997; Nosofsky, 1986).

In natural contexts, people may form a single category, Z, from a stimulus set. In this case, the test of category acquisition is the ability to identify those stimuli as Zs or non Zs. Holyoak (Flanagan, Fried, & Holyoak, 1986; Fried & Holyoak, 1984) found that, after a set of category members was presented, people could make judgments as to category membership of stimuli. If a category is induced from a sample of stimuli drawn from a particular distribution it will have a graded structure; that is, there will be better and worse instances reflecting the decrease in the probability of membership at more extreme values. For certain observed values, only 10%, 5%, or 1% of category instances would be more extreme. Thus, category membership will be uncertain for extreme values within the presented distribution, and also, of course, for values slightly more extreme than those presented. Note that the uncertainty of category membership for extreme values in distributions of equal range will vary with the shape of the distribution.

In acquiring an inductive category, the smallest number of instances people can use to support inferences about the characteristics of the distribution is two. Even this limited information can be used to adjust stimulus values in estimation and to crudely infer boundaries. Because the standard error of the mean, the standard deviation, and any percentile of the distribution of instances is inversely proportional to the square root of the number of instances, the standard error of these summary statistics is cut in half after 8 instances and halved again after 32 instances. In fact, the mean and variance are fairly stable after incorporation of information from 8 to 32 instances. Thus, a person need not experience many instances of a category to have a good sense of its shape.

There are differing views as to how information about a stimulus distribution is preserved in memory—as summary statistics (mean, median, variance, boundaries) or as a set of exemplars from which summary statistics are computed when a category is used in processing. Our model makes only minimal assumptions about representation, and for our purposes, this distinction is not important. We simply assume that summary statistics are available at the time of stimulus estimation. (For a discussion of the indistinguishability of these forms of representation in many cases see Estes, 1986, and Barsalou, 1990.) For the tasks we examine, it seems likely that category information will be preserved in memory as summary statistics rather than as individual exemplars. Stimuli take many values along a dimension (e.g., size, shade of gray). In fact, because the dimensions are continuous the potential number of exemplars is infinite.

Processing

In our model, stimulus estimates are constructed by combining fine-grain and category information. The fine-grain value used in estimation is retrieved from the area of inexactness around the stimulus, based on a process analogous to random sampling (as indicated under I in the Appendix). Establishing the inexactness of that value involves retrieving one or more additional values, again based on a process analogous to random sampling. Here we consider the estimation of stimuli that are treated as members of a particular category (predictions for stimuli at values where membership is uncertain are presented below). A fine-grain stimulus value from the category is adjusted by a weighted mixture with the central category value.[2] The adjustment reduces the variability of estimates although it causes estimates to be biased toward the category center. The *bias* of the estimate of a particular stimulus is the difference between the actual value of the stimulus and its estimated value (as indicated under II in the Appendix).

The more inexact a stimulus representation, the greater will be the bias of estimates of that stimulus. Consider why. At one extreme, if representation is exact, there is no bias because people give the correct value. At the other extreme, if only the category, not the particular stimulus, is remembered, bias is maximal because all estimates are at the category center. Between these extremes, the more exact the stimulus representation, the less the weight that should be given to the central value (as indicated under III.A in the Appendix). Although the bias of estimates should increase monotonically with stimulus inexactness, the variability of estimates should not. When representation is exact, there will be no variability because each estimate will be correct; when only the category is remembered, there also will be no variability because each estimate will be at the central value. Hence variability should first increase and then decrease as stimulus inexactness increases (as indicated under III.B in the Appendix).

Bayesian procedures can always reduce variance enough to more than compensate for the bias introduced, thus increasing accuracy by decreasing the distance (mean-squared error) of an estimate from the true value. Bayes's theorem describes a procedure for computing optimal estimates by trading off bias against variance (as indicated under IV in the Appendix). The mathematical relation of inexactness of stimulus representation to the bias and variability for optimal (Bayesian) stimulus adjustment is shown in Figure 1.

Figure 2A shows a schematic plot of bias for stimuli at different locations across a category. True values are on the horizontal axis, and bias is on the vertical axis. The bias curve shows a downward slope, where estimates for small stimuli are larger than true values and estimates for large stimuli are smaller than true values. The exact shape and slope of the bias curve for a category depends on the extent of inexactness of stimulus representations. Stimulus inexactness is affected by the experimental situation—exposure conditions, interference tasks, and so forth. The more inexact the stimulus representations, the steeper the bias slope should be (as indicated under V in the Appendix). Huttenlocher et al. (1991) showed a steeper bias slope when an interference task was given.

The characteristics of stimuli themselves also affect the inexactness of their representation. For dimensions that

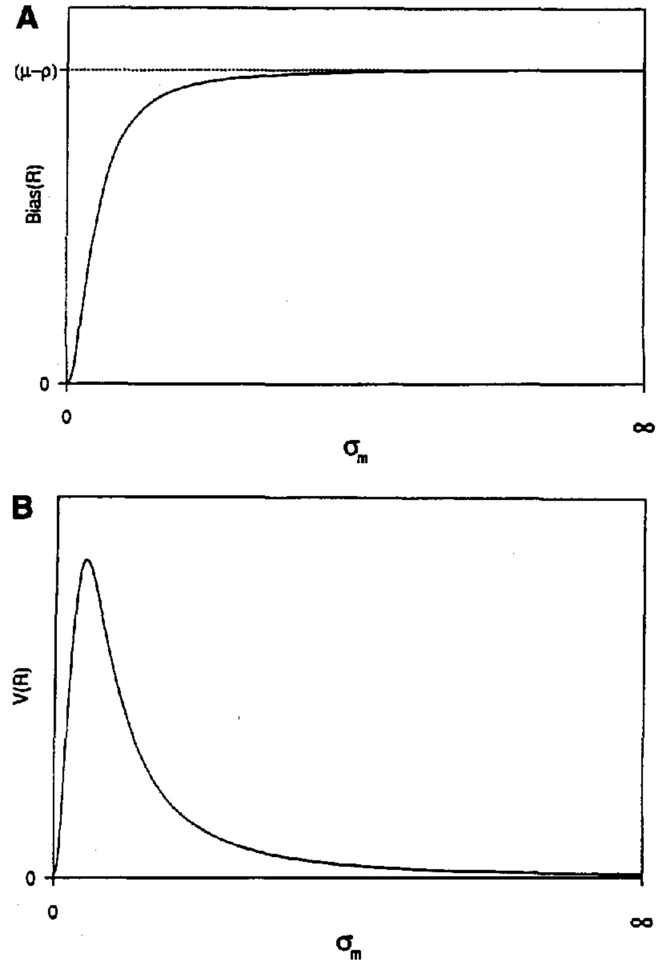

Figure 1. Effects of adjustment using optimal weighting for varying memory inexactness (σ_M) on (A) bias and (B) variability. R = response; V = variability.

increase in magnitude (i.e., where psychophysical scaling is traditionally used), larger values are represented less exactly than smaller values. Therefore, in physical units the bias curve will become steeper where fine-grain values across a category become less exact because the prototype will be more heavily weighted for larger values, as shown in Figure 2B.

Effects of Differences in Prior Distributions on Estimation

This article tests the model's predictions as to how variations in prior (category) information should affect

[2] This is true for an inductive category, consisting of a statistical distribution. In contrast, some categories have exact boundaries (e.g., temporal categories such as academic semesters) such that an inexact stimulus value may be retrieved that lies outside the boundary. In this case, the value may be rejected and another value may be sampled, truncating the distribution of sampled fine-grain values and resulting in large bias near the boundary (see Huttenlocher et al., 1988).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

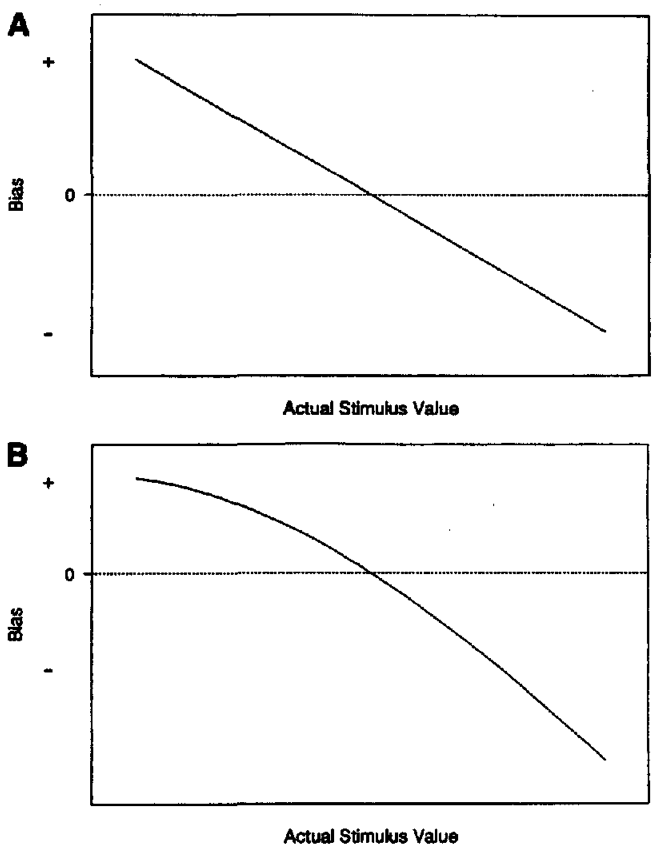

Figure 2. Bias due to adjustment using optimal weighting when memory inexactness of different stimuli is (A) equal and (B) increases with stimulus magnitude.

stimulus estimation. To test the predictions, we contrast judgments for the same stimuli when these are embedded in different distributions. The first prediction is that for distributions with the same range of values, the weight given to category information away from the center should be affected by relative density. At fixed locations toward the ends of the range of values, the probability of membership is higher for uniform than for normal distributions; this should affect the weight given to category information, in turn affecting the bias of estimates for those stimuli. The second prediction is that the concentration of instances in the category should affect the variability of stimulus estimates. In particular, the variability of estimates of all categorized stimuli should be less when the prior distribution (category) is more tightly clustered; this prediction, which follows from our Bayesian model, is not easily derived from other sets of assumptions.

Uncertainty in Categorization

To the extent that an inexactly represented stimulus is judged to be a category member, it should be adjusted toward the center of the category, leading to bias in estimation. To the extent that it is judged not to be a member, it should not be adjusted and hence should be unbiased. Adjustment toward the center of the category increases accuracy for a category member and decreases accuracy otherwise. It seems intuitively reasonable to use an adjustment of stimuli proportional to the judged probability of category membership (as indicated under VI.A in the Appendix). In fact, proportional adjustment, precisely stated, can be shown to be a Bayesian estimate and therefore optimal (as indicated under VI.B in the Appendix).

The exact weight given to the category in adjusting extreme stimulus values depends, as we noted above, on the shape of the distribution; it also depends on the consequences of error (e.g., given the consequences of wrongly considering a poisonous mushroom to be edible, people may be expected to be more inclusive in what they treat as poisonous than in what they treat as edible). The issue of the consequences of error raises a broad set of issues that is beyond the scope of the present model (although the issue is discussed briefly at the end of this article). The prediction we make here is that if the consequences of error are equal for two categories of the same range, the category weights at particular locations should be affected by the shape of the distribution (i.e., the relative density of extreme values). The relative density of instances at locations away from the center of the category is lower for a normal distribution than for a uniform distribution. Thus, when the consequences of error are equal, the weight given to the category should decrease more rapidly at extreme locations for a normal than for a uniform distribution. Hence the bias function should show higher curvature for the normal, flattening out and then decreasing toward zero as the probability that a stimulus is a category member falls to zero.

Finally, note that the likelihood that a stimulus in a fixed position in a given distribution will be judged to be a category member should be affected by the inexactness of its representation. The less exact a stimulus representation is, the greater is the range of values that may be retrieved for that stimulus. For scales of increasing magnitude where stimulus inexactness increases with magnitude, there are two opposing effects. First, as we saw above, in regions where category certainty is high, the bias slope will increase as magnitude increases because the prototype is weighted more heavily for more inexact values. Second, at a fixed value that has a very low probability of membership, far from the category center, bias will be less when the stimulus is less exactly represented because the area of stimulus inexactness extends farther from the category center.

Variability of Instances

A stimulus that is judged to be a category member will be adjusted toward the center. There are two consequences of stimulus adjustment: It reduces the variability of estimates and it introduces bias. A Bayesian model predicts that the variability of estimates of stimuli, at all locations in that category, will be affected by the distribution. In particular, estimates of a categorized stimulus from a category where instances are more concentrated toward the center should be less variable than estimates for the same stimuli embedded in a category where instances are more dispersed. Consider why. At one extreme, if a category has just one stimulus value, the variability of estimates of that stimulus will be zero. At the other extreme, if instances are very dispersed, the variability of estimates will be maximal because the central category value provides minimal information about a particular stimulus, and hence retrieved values will be adjusted negligibly. Between these extremes, the greater the concentration of instances is, the smaller is the variability of estimates (see VII.A in the Appendix). Hence, variability of the estimates of categorized stimuli should be less for a normal than a uniform distribution, and variation should be less for distributions of the same shape when the range of values is narrower.

Note that there is not a monotonic relation between the concentration of instances in the category and the bias of estimates. At one extreme, if a category has only one value, bias is necessarily zero. At the other extreme, if the concentration of instances is very low, adjustment will contribute minimally to accuracy, so again there should be little bias. Between these extremes of concentration, bias should first increase and then decrease (see VII.B in the Appendix). The relation of concentration of instances to variability and bias for optimally weighted combinations of fine-grain and prototype information is shown in Figure 3.

The Experiments

In each of the three experiments in this article, we taught people a category by presenting a set of stimuli that varied along a single dimension, forming one cluster over a range of values. Stimuli were presented one at a time on a computer screen. Participants reproduced each stimulus after it had been removed from the screen. We measured the bias of the estimates for each stimulus; we expected to find a single bias curve, providing evidence that people form a single category. We did pilot work preparatory for the study in which people first saw a series of stimuli (fish that varied in fatness) and then judged category membership for a new series of stimuli where the range of values was extended to include stimuli outside the range of the initial set. People were very accurate over the center range of values and, even near extremes, were more likely to judge stimuli from the set previously seen to be members than stimuli from outside that range. The probability of judging extreme values to be members fell off more rapidly for normal than uniform distributions of the same range.

In each of our experiments we manipulated the distribution of stimuli; different groups of participants were presented with different distributions. One group was given a uniform distribution, another group was given a normal distribution of the same range of presented stimuli as in the uniform condition, and two other groups were given narrower uniform distributions, comprising either the lower or upper half of the stimulus values.[3] The distributions are shown in Figure 4. We examined reproductions as a function of actual stimulus size for normal and uniform distributions. Our model predicted that the pattern of bias across the category would be nonlinear for the normal condition because the uncertainty as to membership decreases the

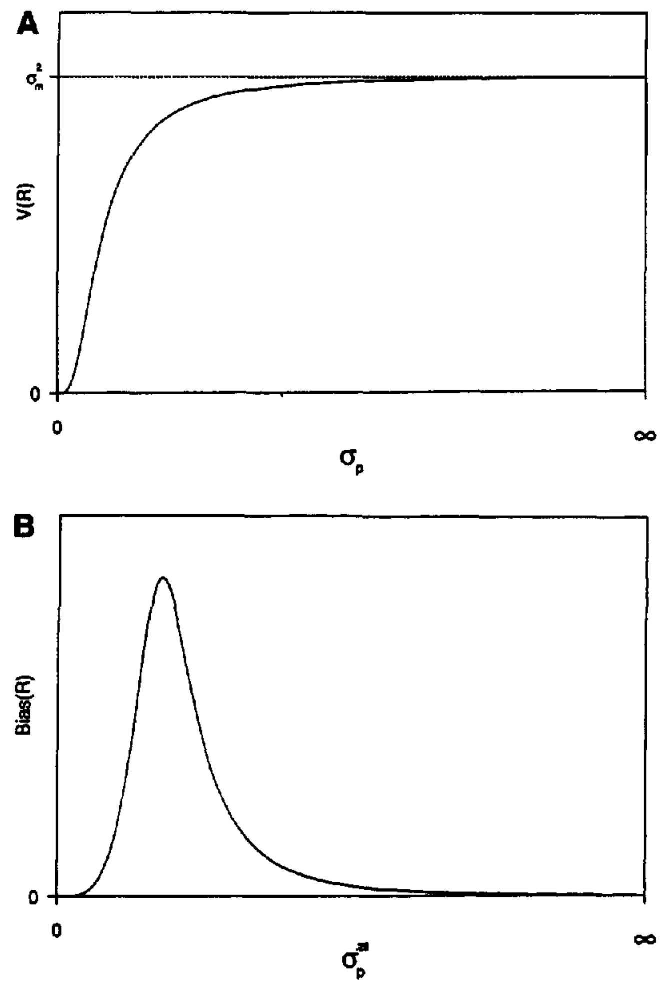

Figure 3. Effects of adjustment using optimal weighting for varying category concentration (σ_P) on (A) variability and (B) bias. R = response; V = variability.

extent of adjustment toward extreme values. We also evaluated the effects of the dispersion of instances, comparing the variability of stimulus estimates for the wide uniform distribution with the two narrow uniform distributions and the normal distribution. Our model predicted that the standard deviation of estimates of stimuli would be greatest for the wide uniform distribution.

We used three different dimensions in our three experiments. For each of the dimensions, distances along the

[3] Note that the distribution of stimuli presented in the "normal" condition of our experiments is not precisely a normal distribution. Rather, it is something of a hybrid between a normal and a triangular distribution. For our purposes, the essential characteristics of the distribution are that it is peaked in the center and sparse in the tails and that it contains the same range of stimulus values that are presented in the uniform stimulus condition. We continue to refer to the distribution as "normal" only to avoid the necessity of repeating a phrase such as "the distribution that embodies the characteristics of a normal distribution that are important in our model."

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

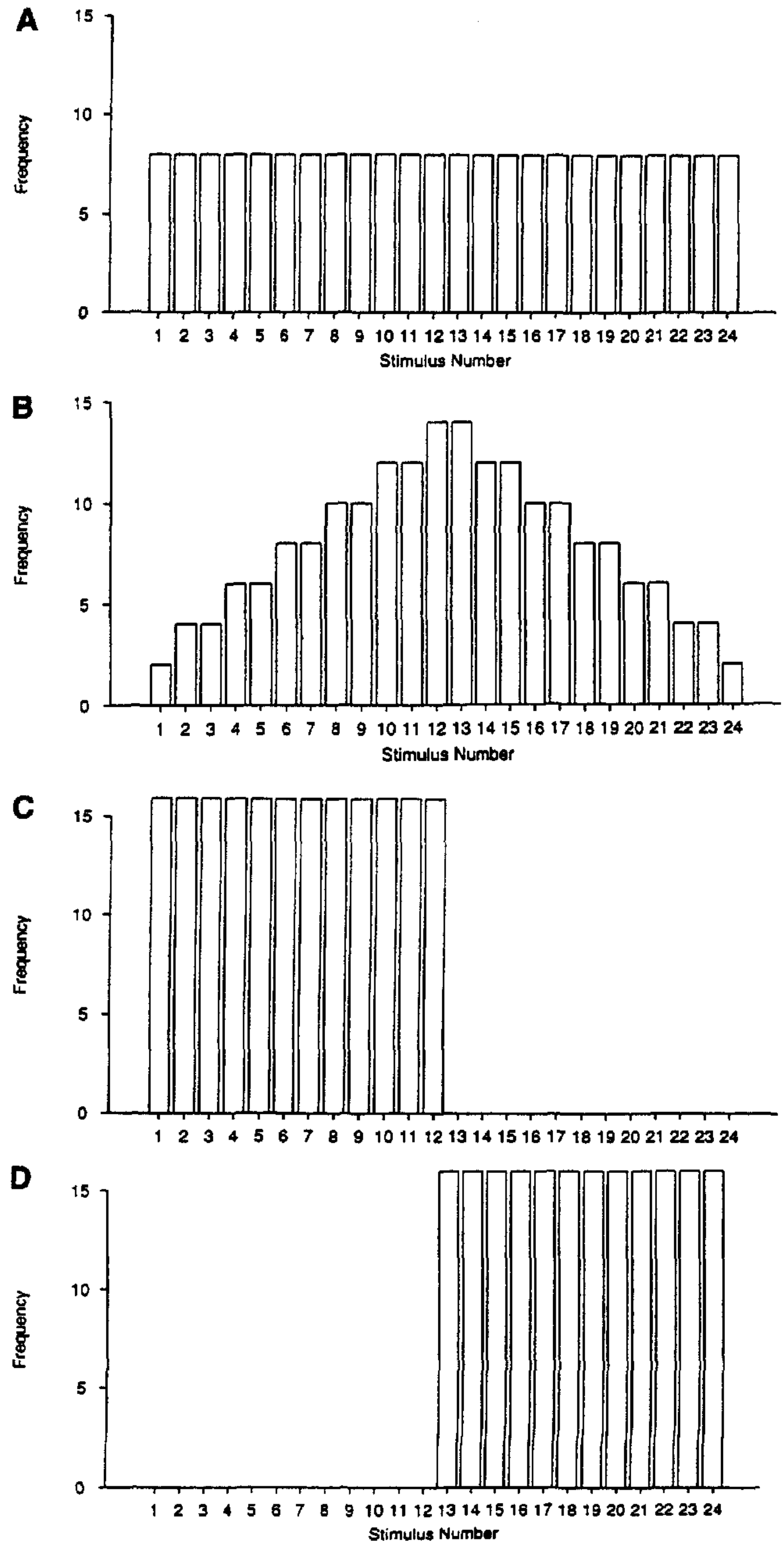

Figure 4. The distributions of stimuli presented in the experiments: (A) uniform distribution, (B) normal distribution, (C) narrow uniform distributions (small), and (D) narrow uniform distribution (large).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

dimension could be measured in equal size physical units—fatness of fish, grayness of squares, and lengths of lines. However, the dimensions could be processed differently, and we used the three to test whether the model holds fairly generally. The dimension of fatness had zero as an absolute boundary at the slender end; general principles could be used to impose an upper bound (e.g., that a fish isn't fatter than it is long). Shade of gray was bounded at both ends—black at one extreme and white at the other; it was unclear whether shade would be treated as a scale that increases in magnitude. The dimension of length clearly increased in magnitude, and it is one where psychological scaling generally is used. For length, then, the inexactness of representation should have increased with the magnitude of the stimuli. This should have had two consequences. First, the bias slope should have accelerated because the central value was given increasing weight as stimulus inexactness increased. Second, the bias slope should have leveled off and begun to decrease sooner because when stimuli are less exactly represented particular instances retrieved from memory will be more likely to fall far from the category center.

Experiment 1: Fish

In this study, participants were shown fish that varied in fatness. After each fish was presented and removed, participants reproduced its size. We examined the estimation of stimulus values in four different groups of participants who were presented with one of the four stimulus distributions shown in Figure 4. We assessed the prediction that bias would fall off nearer the category center for a normal distribution than for a uniform distribution of the same range. We also assessed the prediction that variability of estimates would be less for narrower categories and for the normal distribution in the region where certainty of membership is high.

Method

Participants. The participants were undergraduate and graduate students at the University of Chicago drawn from a list of people interested in participating in psychology experiments. There were 10 participants in each of the four conditions. They were paid $5 for a session lasting 30 min.

Materials. The presentation of stimuli and the collection of responses were controlled by a program that ran on a Macintosh computer connected to a large (38 cm wide × 29 cm tall) monochrome monitor with a resolution of 30 pixels per cm. (Pixels are used to describe the stimulus and response values because they are the incremental unit used to adjust the images.) The stimuli were figures representing fish, consisting of an elliptical body, a fan-shaped tail, and a round eye. They varied only in fatness (the vertical dimension). The other dimensions of the fish were constant: The body was 400 pixels long (13.33 cm), the eye was 30 pixels in diameter (1 cm), the tail was 240 pixels (8 cm) long, and the back of the tail swept in 120 pixels (4 cm). The height of the tail was proportional to the height of the body but 10% smaller. The body and tail of the fish were light gray, and the eye was black. The fish were presented in the center of the computer monitor (see Figure 5 for a scaled-down depiction of a typical fish within the boundaries of the monitor).

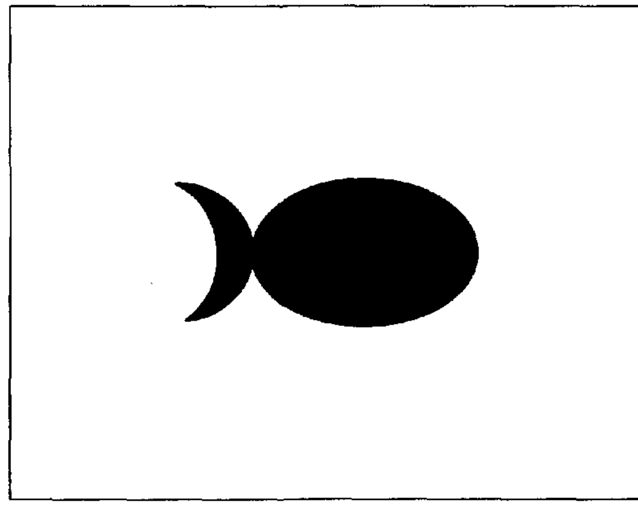

Figure 5. Illustration of the schematic fish stimuli used in Experiment 1.

In both the uniform and normal conditions there were 24 distinct stimuli varying in body height from 160 pixels (5.33 cm) to 344 pixels (11.47 cm) in increments of 8 pixels (0.27 cm). In the uniform condition, each stimulus was presented once within each of eight blocks for a total of 192 trials. In the normal condition, stimuli near the mean of the distribution were presented more frequently than stimuli near the endpoints of the range. Within a block, the distribution of stimuli was as follows: The two most extreme stimuli (160 and 344 pixels) were shown once; the next two stimuli from the tails of the distribution (168, 176, 328, and 336 pixels) were shown twice; stimuli with a fatness of 184, 192, 312, and 320 pixels were shown three times; stimuli with a fatness of 200, 208, 296, and 304 pixels were shown four times; stimuli with a fatness of 216, 224, 280, and 288 pixels were shown five times; stimuli with a fatness of 232, 240, 264, and 272 pixels were shown six times; and the most central stimuli (248 and 256 pixels) were shown seven times. There were 96 trials in each of two blocks for a total of 192 trials. In the two narrow conditions, a smaller range of stimulus values was presented. In the slender half, the fatness of the fish varied from 160 to 248 pixels. In the fat half, the fatness varied from 256 to 344 pixels. In these two conditions, there were 12 stimuli presented once in each of 16 blocks. There were 192 trials in all four conditions.

Procedure. On every trial of the experiment, a fish was shown in the center of the monitor for 2 s. Then, the screen went black for 200 ms, and after a delay of 1 s, the participants adjusted the size of a response fish, attempting to match the size of the fish they had just seen. The response fish had an initial height of 100 pixels (13.33 cm). To prevent participants from using landmarks to estimate the height of the fish, we placed the response fish 65 pixels (2.16 cm) lower than the stimulus fish and 104 pixels (3.46 cm) to the right. The participants adjusted the height of the response fish by pressing the "L" key on the keypad to make the fish larger and pressing the "S" key to make the fish smaller. When the participants were satisfied with their estimate, they pressed the spacebar. Finally, the participants heard a beep, which signaled them to prepare for the next trial.

Scoring. On some trials, the participants did not make a response. Instead of pressing the "S" and "L" keys to adjust the fatness of the initial fish, they pressed the spacebar by mistake. To prevent these cases from affecting the mean error, we deleted

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

responses that exactly equaled the initial fatness of the fish. Nonresponses occurred on only 0.99% of the trials in the uniform condition, 1.21% in the normal, 0.94% in the slender half, and 0.89% in the fat half. We determined bias in estimation by subtracting the actual fatness of the fish from the fatness of the participant's response. Some of the biases were so large that it seemed the participant must have accidentally entered an arbitrary response, perhaps because of striking the spacebar prematurely or momentarily failing to attend to the task. To detect these cases, we calculated quartiles of the distribution of responses for each stimulus value, and we deleted responses deviating from the median by more than three interquartile ranges (IQRs). The number of responses deleted was relatively small. Both culling procedures removed a total of 1.51% in the uniform condition, 1.56% of responses in the normal condition, 1.09% in the slender half, and 1.25% in the fat half.

Results

Pattern of bias. For each of the four conditions in Experiment 1, the mean bias is plotted as a function of the actual fatness of the stimulus (see Figure 6). Under all conditions, the bias ranges from positive for slender fish to negative for fat fish and crosses zero near the center of the distribution of stimuli. Such a pattern of bias supports the prediction that estimates are shrunken toward a central value. That is, participants tend to overestimate the fatness of more slender fish in the presented distribution and underestimate the fatness of fish that are fat relative to the distribution. Thus, in the normal and uniform conditions, the bias is zero in the vicinity of 252 pixels; for the slender condition, the bias is zero near 204 pixels; and for the fat condition, the bias is zero near 300 pixels.

Bias shape. The model predicts that for categories that are equal in range the shape of the bias function is affected by whether the distribution of presented stimuli is normal or uniform. In the normal condition there should be less bias near the boundaries because extreme stimulus values are less likely to be considered category members than in the uniform condition. As can be seen in Figure 6, the prediction appears to be correct. The difference between the curves for the distributions is substantial. At the slender end of the distribution the bias for the normal condition is flat, whereas the bias for the uniform condition is linear. Similarly, at the fat end the bias curve for the normal condition flattens out, whereas the curve for the uniform condition continues its linear pattern. This again is consistent with the prediction that boundary uncertainty is greater in the normal condition.

In order to determine whether the shape of the bias curve in the normal and uniform conditions is different, it is desirable to compare a numerical index of bias shape. Because decreased bias at the ends of the range results in a less linear bias function, we examined the degree to which bias departed from linearity. One such index of linearity is obtained by considering the problem as a special case of repeated measures and using orthogonal polynomial contrasts. For the uniform and normal conditions, we performed an analysis of variance using each participant's mean response to each stimulus value and treating the stimuli as different levels of a factor repeated within participants. A sum of squares was calculated for each orthogonal polyno-

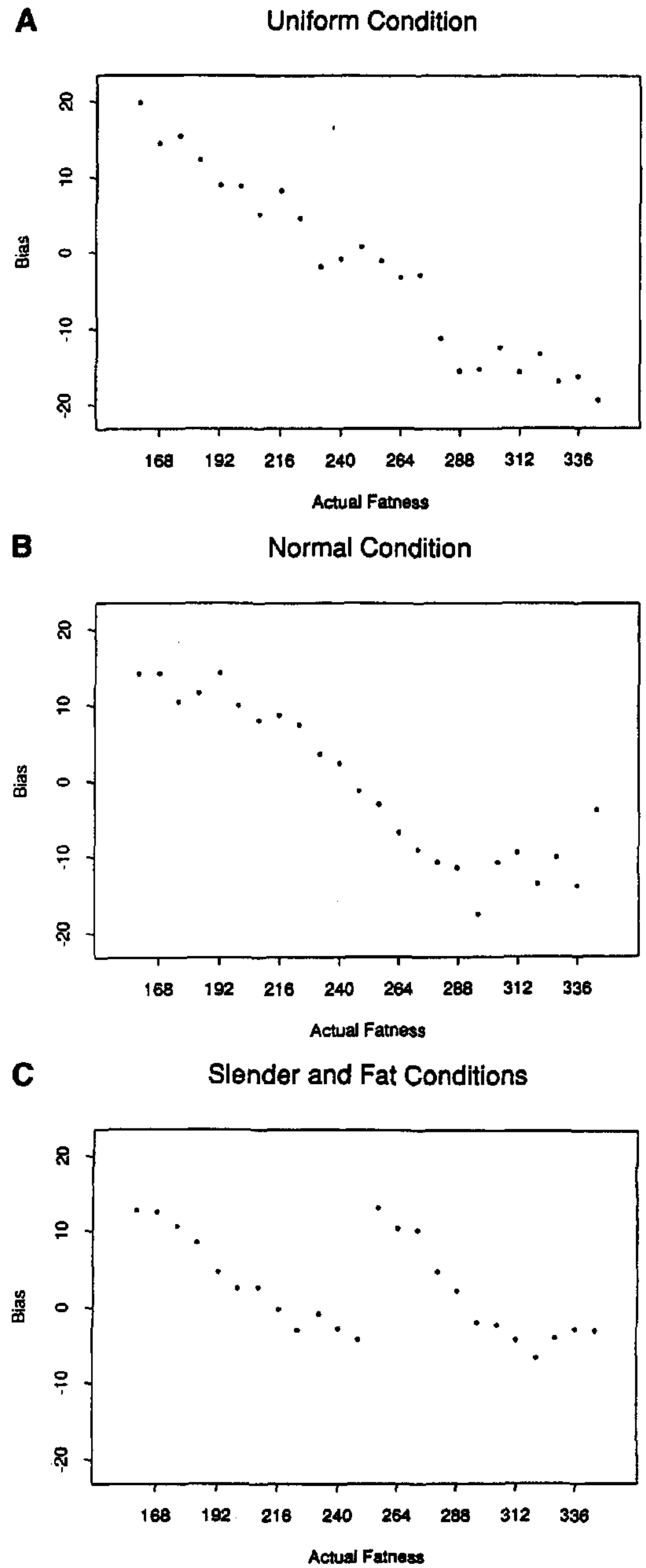

Figure 6. Mean response bias in reported fish size, plotted as a function of actual fish size under (A) the uniform condition, (B) the normal condition, and (C) the slender and fat conditions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

mial component; the degree to which the linear component accounts for total variability was estimated by η^2, the ratio of the linear component's sum of squares to the total of the linear and nonlinear components' sums of squares.

Values of η^2 approaching 1.0 indicate that the bias pattern is primarily linear, whereas lower values indicate departures from linearity. In the present experiment, we expect a high value for the uniform condition and a lower value for the normal condition. The values actually obtained were .953 for the uniform condition and .827 for the normal condition. That is, the linear trend accounts for about 95% of total explained variability under the uniform condition compared with only about 83% for the normal condition. The difference may be attributed to the flattening of the bias curve at the slender end and the hook toward zero at the fat end under the normal condition.

A test for the difference between the two values of η^2 is not available, so we developed an ad hoc procedure that is equivalent to such a test in the critical respects. We combined the normal and uniform data, incorporating condition (uniform vs. normal) as a between-groups factor. We also included in the model the orthogonal polynomial contrasts on stimulus magnitude, as well as terms representing the interaction between contrast components and stimulus condition. We weighted the analysis to compensate for heteroscedasticity resulting from analyzing means based on different sample sizes. We then conducted a nested F test on the collective sum of squares of the condition-by-contrast interaction for the quadratic and cubic terms. (For practical purposes, the linear and quadratic components account for the bulk of the observed curvature.) That is, we observed the change in the model sum of squares when those two interactions were dropped from the model, transformed the sum of squares for change into a mean square based on two degrees of freedom, and divided by the error term for repeated measures effects. The resulting $F(2, 414)$ was $4.597, p = .011$. The implication is that the greater curvature associated with the lower value of η^2 in the normal condition is statistically reliable.

Standard deviation of bias. The model predicts that the standard deviations of stimulus estimates should vary for different stimulus distributions. First, responses for the normal condition should be less variable than responses for the uniform condition. Second, responses for a uniform distribution should be less variable when the range of stimuli is smaller (as when the narrow slender and fat uniform conditions are compared with the wide uniform condition).

First, let us compare the standard deviations for the normal and uniform distributions. We should restrict ourselves to a region within the categories where the certainty of membership is equal. Because the certainty that a stimulus is in the category decreases more markedly near the boundaries for a normal distribution than for a uniform distribution, we elected to compare standard deviations over a central region where participants were quite certain of the category for both distributions. Hence, we focused our attention on the 10 most central stimuli. For each of those stimuli, we calculated the standard deviation of raw responses. Then, we calculated the natural logarithm of those standard deviations and compared the mean log standard deviation for the uniform and the normal conditions using the two-sample independent groups t test. (The log transformation is the appropriate variance stabilizing transformations; see Bartlett & Kendall, 1946. Results did not differ appreciably for untransformed data.) The mean log standard deviation for the normal condition was 3.003; for the uniform condition, the value was 3.177. The difference is highly significant, $t(18) = 4.875, p < .0002$. (See Table 1 for a summary of the standard deviation comparisons for both logged and untransformed values.)

Next, let us compare the standard deviations for uniform distributions that vary in width. Because the shapes of the distributions were the same, the issue of category uncertainty seen in the comparison of normal and uniform distributions was not a concern. Thus, we compared the logged standard deviations for all values in the slender and fat half distributions with the comparable values in the wide uniform distribution. The average transformed standard deviation in the slender half distribution was 2.808, and the average for the slender half of the wide distribution was 3.116. The difference was highly significant, $t(22) = 8.057$, $p < .00001$. Note that participants in the slender half condition saw twice as many examples of each stimulus as did participants in the wide uniform condition; that could provide an alternate explanation for observed differences in standard deviations. Accordingly, we repeated the test using only the first half of the slender condition trials. The results were virtually identical: The average logged standard deviation was 2.814, and the test for the difference between that value and the average for the narrow half of the wide uniform condition remained highly significant, $t(22) = 8.545$, $p < .00001$. We also compared average logged standard deviations from the fat half condition with the corresponding standard deviations in the wide uniform condition. The mean transformed value for the fat half trials was 2.977 (or 2.964 when only the first half of the trials was considered). The mean logged standard deviation for the fat half of the wide uniform condition was 3.285. The difference was highly significant, $t(22) = 7.981, p < .00001$; $t(22) = 7.826, p < .00001$, when only the first half of the fat trials was considered. The variability of responses, then, was lower for every condition in which information about the category was more precise.

Table 1
Variability in Experiment 1

Condition	*SD* (in pixels)	Log *SD*	*t*	*df*	*p*
Normal	19.950	3.003			
Uniform	24.033	3.177			
Difference	4.083	0.174	4.875	18	.00012
Uniform	22.611	3.116			
Skinny half	16.678	2.808			
Difference	5.933	0.308	8.057	18	5.246×10^{-8}
Uniform	26.870	3.285			
Fat half	19.673	2.977			
Difference	7.197	0.308	7.981	22	6.141×10^{-8}

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Discussion

This experiment shows that categories were used in estimating stimulus values, with responses biased toward a central value (prototype) corresponding to the mean or median of the presented stimuli in all four conditions. The prototype was located at the same value for the uniform and normal conditions, where the range of presented stimuli was the same, and at a different value in the slender and fat half conditions where the range differed. The findings supported the predictions of the model.

First, the shape of the bias function was affected by the shape of the distribution for distributions of the same width. According to the model, the extent of the bias is affected by the probability that stimuli will be included in the category. When instances are normally distributed, there are fewer stimulus values near the tails of the distribution, which should affect the subjective certainty of category membership and hence prototype weighting, leading to less bias for these stimuli. As predicted, these extreme stimulus values showed less bias in the normal condition than in the uniform condition.

Second, the variability of responses was less when the concentration of instances was greater. According to the model, variability of estimates should be less when a category prototype is more precise. The precision of the prototype was increased by reducing the dispersion of instances about a central value in the normal versus the uniform condition and by decreasing the range of presented instances in the split half condition. In both of these cases, the variability of responses was less than that in the wide uniform condition.

Experiment 2: Shades of Gray

To test our model of the processes involved in estimating stimuli from memory, it is important for us to show that the same pattern of results holds for different stimulus dimensions when the type of prior distribution is varied. Hence, we examine a very different sort of dimension, shade of gray, to determine whether people form categories and whether the categories formed include information as to the dispersion of instances.

Method

Participants. The participants, undergraduate and graduate students at the University of Chicago, were drawn from a list of people interested in participating in psychology experiments. There were 10 participants in each of the four conditions. They were paid $5 for a session lasting 30 min.

Materials. The stimuli were squares that varied in shade of gray. The squares were 200 pixels (6.67 cm) in size and presented in the center of a 33-cm computer monitor. The room was dimly lit by a single tungsten bulb. The shades of gray are described in terms of the number of photometer units emitted by the computer monitor. In the uniform and normal condition, there were 24 stimuli varying in shade of gray from 700 (light) to 4,150 (dark) photo units in increments of 150. (The units represent a linear transformation of photometer readings taken directly from the computer screen, resulting in a scale of darkness.) The stimuli in the uniform condition were presented once in each of 8 blocks for a total of 192 trials. In the normal condition, stimuli near the mean of the distribution were presented more frequently than stimuli near the endpoints of the range. Within each block, the distribution of stimuli was as follows: once at 700 and 4,150; twice at 850 and 4,000; three times at 1,000, 1,150, 3,700, and 3,850; four times at 1,300, 1,450, 3,400, and 3,550; five times at 1,600, 1,750, 3,100, and 3,250; six times at 1,900, 2,050, 2,800, and 2,950; and seven times at 2,200, 2,350, 2,500, and 2,650. There were 96 trials in each of 2 blocks for a total of 192 trials. In the light half condition, the shade of gray varied from 700 to 2,350. In the dark half condition, the shade varied from 2,500 to 4,150. In these two conditions, there were 12 stimuli presented once in each of 16 blocks.

Procedure. On every trial, a gray square was shown in the center of the monitor for 2 s. Then, the screen went black for 200 ms. After a 1-s delay, the participants adjusted the shade of a response square to match the shade of gray they had just seen. The response square had an initial shade that was lighter than any stimulus. The participants adjusted the shade of gray by pressing the "L" key on the keypad to make the square lighter and the "D" key to make the square darker. When the participants were satisfied with their estimate, they pressed the spacebar. Then, the participants heard a beep, which signaled them to prepare for the next trial.

Scoring. On some of the trials, the participants pressed the spacebar by mistake instead of pressing the "L" and "D" keys to adjust the shade of the initial gray square. These nonresponses occurred on only 0.05% of the trials in the uniform condition, 0.16% in the normal, 0.36% in the dark half, and 0.16% in the light half, and they were deleted. Error in estimating the shade of gray was determined by subtracting the actual shade of gray from the shade of the participant's response. To detect cases where the participant seemed to have responded arbitrarily, we deleted responses that deviated from the median by more than three IQRs. The culls removed a small portion of the trials, 0.31% in the uniform condition, 0.21% in the normal, 0.63% in the dark half, and 1.46% in the light half. Finally, the mean error for each shade of gray was calculated from the remaining responses.

Results

Pattern of bias. In Figure 7, the bias is plotted against stimulus magnitude for the uniform, normal, dark, and light gray conditions. Once again, the bias patterns are consistent with the prediction that responses are shrunken toward a prototype. The bias ranges from positive for light grays to negative for dark grays and is zero near the center of distribution of presented shades. There is some suggestion of a hook toward zero bias in each of the four conditions; the hook appears for stimulus values at the dark end of the range and is more pronounced in the normal than in the uniform condition of the same range. The hooks are less pronounced for the grays than for the fish, perhaps because the gray stimuli covered such a large portion of the possible values that there is no possibility that other categories of gray might lie past the range encompassed by the shades presented. Hence, the flattening of the bias curve is probably partly attributable to increasing scale magnitude as stimuli approach the dark end of the continuum.

Bias shape. Again, the model predicts that the uniform condition should exhibit a more linear bias pattern than the normal condition. The index η^2 supports that prediction. For the uniform condition, η^2 is equal to .904. For the normal

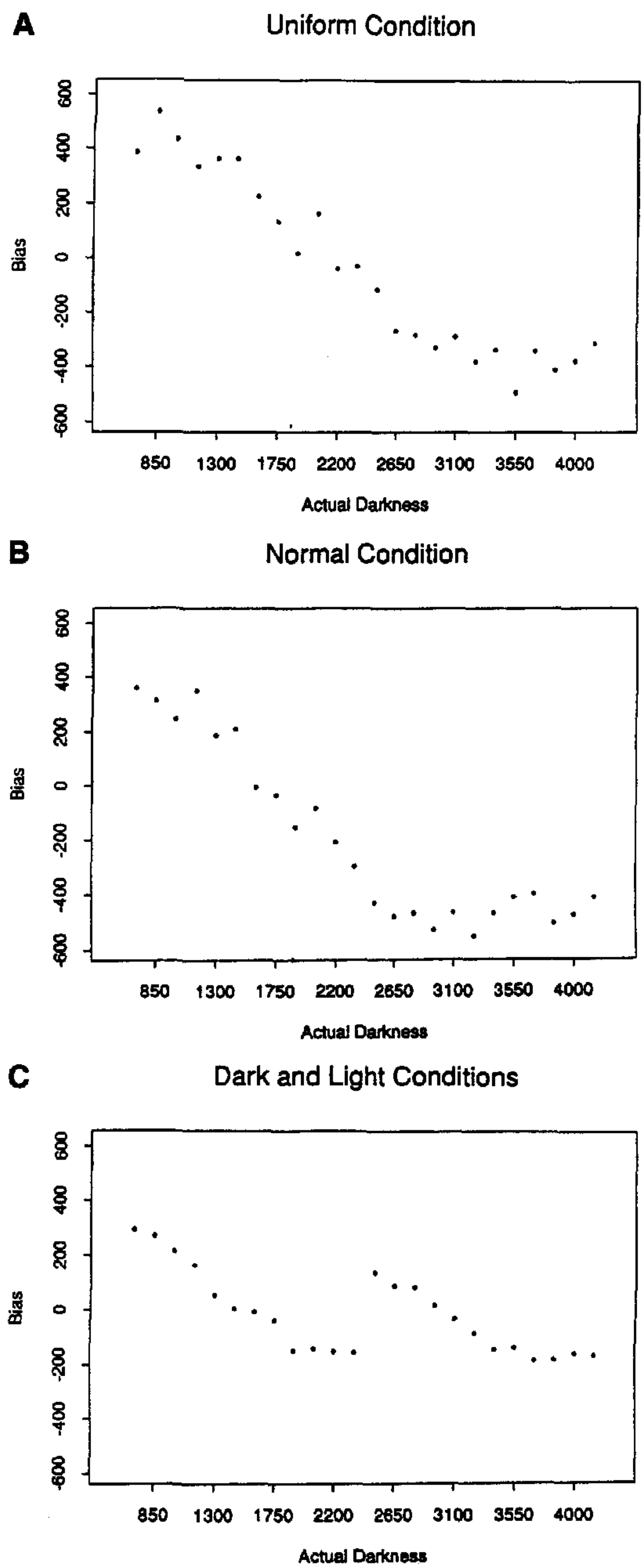

Figure 7. Mean response bias in reported shade of gray, plotted as a function of actual shade under (A) the uniform condition, (B) the normal condition, and (C) the dark and light conditions.

condition, it is .824. That is, about 90% of the explained variability in the uniform condition is accounted for by the linear trend, compared with only 82% for the normal condition. We conducted the same nested F test for the difference in curvature that we described in Experiment 1. The resulting $F(2, 414)$ had the value 0.015, $p = .985$, *ns*. Thus, although the indices of curvature are consistent with the proposition that the normal condition has a more curvilinear bias pattern, any difference between the normal and uniform conditions is shrouded in such a high degree of noise that it is statistically trivial. This is not unexpected, because the very nature of the gray stimuli is such that there are ultimate bounds (black and white) beyond which stimuli cannot exist. Hence, there is a natural limit to any curvature effects associated with a tendency for extreme stimuli in the normal condition to be perceived as outside the category.

Standard deviation of bias. The model's prediction is that the standard deviations of stimulus estimates should vary for different stimulus distributions. Responses for the normal condition should be less variable than responses for the uniform condition, and responses for the dark and light half uniform distributions of stimuli should be less variable than responses for the wide uniform condition. Those predictions were tested in the same manner as in Experiment 1. That is, logs of the standard deviations in the normal and uniform conditions were compared for the central 10 stimuli. The average logged standard deviation for the normal distribution was 6.801; the average for the uniform condition was 6.866. The difference, while in the predicted direction, was nonsignificant, $t(18) = 1.509$, $p < .15$. The standard deviations for the light and dark half distributions were compared with those of the corresponding stimuli in the wide uniform distribution. The average logged standard deviation in the light half distribution was 6.584, and the value for the corresponding stimuli in the wide uniform condition was 6.870. The difference was highly significant, $t(22) = 7.369$, $p < .00001$, and the result was not appreciably different when only the first half of the light presentations was considered, mean logged $SD = 6.568$, $t(22) = 6.678$, $p < .00001$. The average log-transformed standard deviation in the dark distribution was 6.568, and the average for the corresponding stimuli in the wide uniform condition was 6.839. The difference was highly significant, $t(22) = 5.960$, $p < .00001$. The difference remained significant when only the first half of the dark presentations was considered; in that case, the mean logged standard deviation for the dark distribution was 6.536, $t(22) = 5.440$, $p < .00002$. Thus the variability of responses was lower for each condition in which information about the category was more precise, although the difference failed to reach significance in the comparison of the uniform and the normal distributions. The differences are summarized in Table 2.

Discussion

As in the previous experiment, bias toward a prototype located at the mean of the presented stimuli occurred in all conditions. There was not a significant effect of the distribu-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Variability in Experiment 2

Condition	*SD* in Pixels	Log *SD*	*t*	*df*	*p*
Normal	902.452	6.801			
Uniform	963.326	6.866			
Difference	60.874	0.065	1.509	18	.1487
Uniform	966.081	6.870			
Light half	726.700	6.584			
Difference	239.381	0.286	7.369	18	2.246×10^{-7}
Uniform	937.494	6.839			
Dark half	716.580	6.568			
Difference	220.914	0.271	5.960	22	5.350×10^{-6}

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

tion of stimuli within a category on the shape of the bias curve. This may be because the grays presented in the uniform and normal conditions ranged from nearly white to nearly black. Thus, participants may have been more likely to include all of the presented stimuli in the category because other categories of gray were not possible. Also, the variability of responses was affected by the distribution of presented stimuli. There was less variability of responses when the dispersion of instances about a central value was less. Parallel to Experiment 1, variability of responses was substantially greater for the wide uniform than for the half uniform distributions. The difference in variability for normal versus uniform distributions was in the predicted direction, although it was not significant.

Experiment 3: Line Length

To test our model of the processes involved in estimating stimuli from memory, we extend the set of dimensions for which we examine category concentration to lines that vary in length.

Method

Participants. The participants, undergraduate and graduate students at the University of Chicago, were drawn from a list of people interested in participating in psychology experiments. There were 10 participants in each of the four conditions. They were paid $5 for a session lasting 30 min.

Materials. The experiment was conducted on a Macintosh computer connected to a small color monitor 650 pixels by 500 pixels (21.67 cm × 16.67 cm). For uniform and normal conditions, the stimuli were 24 black lines varying in length from 45 pixels (1.5 cm) to 390 pixels (13 cm) in 15 pixel (0.5 cm) increments. In the uniform condition, the lines were presented once in each of 8 blocks for a total of 192 trials. In the normal conditions, the distribution of stimuli within each block was as follows: once at 45 and 390; twice at 60 and 375; three times at 75, 90, 345, and 360; four times at 105, 120, 315, and 330; five times at 135, 150, 285, and 300; six times at 165, 180, 255, and 270; and seven times at 195, 210, 225, and 240. There were 96 trials in each of 2 blocks for a total of 192 trials. For the short half distribution, the lines varied from 45 to 210 pixels. For the long half distribution, the lines varied from 225 to 390. For both of these half distributions, the 12 lines were presented once in each of 12 blocks for a total of 144 trials. The height of the line was always at 7 pixels (0.23 cm).

Procedure. The participants' task was to reproduce the length of the horizontal line segment they saw. On every trial, a line was presented in the center of the monitor for 2 s. Then, the screen went black for 200 ms and, after a 1-s delay, the participants adjusted a response line to match the line that had just been presented. The response line had an initial length of 2 pixels. The participants pressed the "S" key to make the line shorter and the "L" key to make the line longer. When the participants were satisfied with their estimate, they pressed the spacebar. The screen was cleared, there was a brief delay, and the next trial was presented.

Scoring. As in the previous experiments, there were some nonresponses where the participants pressed the spacebar by mistake. Thus, responses were deleted whenever they exactly equaled the initial length of the line. Nonresponses occurred on only 0.52% in the uniform condition, 0.36% in the normal, 0.28% in the short half, and 0.21% in the long half. Error in estimation was determined by subtracting the actual length of the line from the length of the participant's response. We deleted responses that fell more than three IQRs from the median of each stimulus; as in the previous experiments, it seemed that such responses must be the result of participants' momentary failure to attend to the task. The culls removed a small portion of the trials, 0.63% in the uniform condition, 0.36% in the normal, 0.63% in the short half, and 0.34% in the long half. Finally, the mean error for each stimulus length was calculated from the remaining responses.

Results

Pattern of bias. In Figure 8, the mean bias is plotted against stimulus magnitude for each condition. For all conditions, responses are shrunken toward a central value. There is overestimation of short line lengths and underestimation of long lengths.

Bias shape. The prediction that the pattern of bias should be less linear for the normal condition than for the uniform was supported. For the normal condition, η^2 is equal to .724; for the uniform condition, the value is .894. That is, 89% of the explained variability in the uniform condition is accounted for by the linear trend, compared with only 72% for the normal condition. Comparing the normal and wide uniform conditions, we note that there is a more pronounced hook toward zero bias for the largest stimuli in the normal condition. It should be noted that, contrary to fatness of fish, the bias curve at the narrow end flattens even for a uniform distribution. Moreover, the bias curve shows curvilinearity consistent with the increasing stimulus magni-

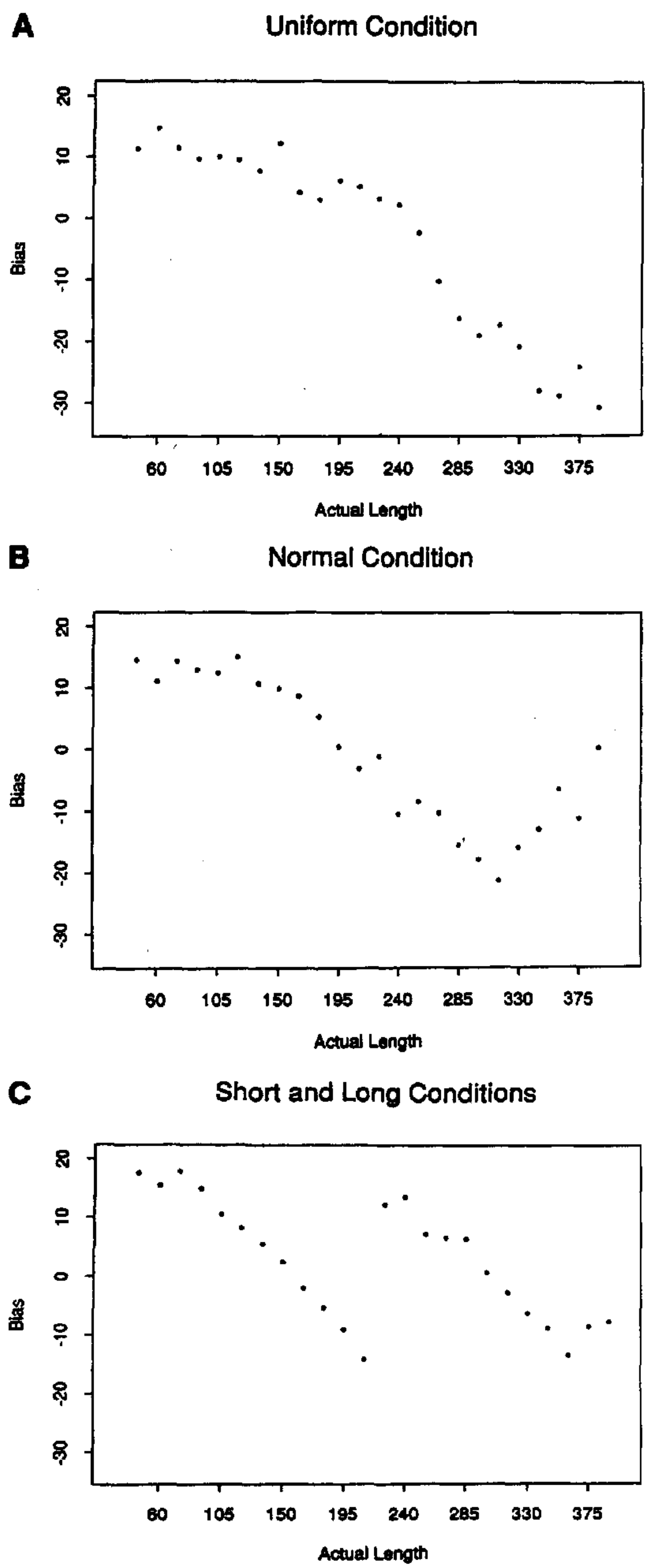

Figure 8. Mean response bias in reported length of line, plotted as a function of actual line length under (A) the uniform condition, (B) the normal condition, and (C) the short and long conditions.

tude. That is, short line lengths appear to have more exact fine-grain representation, as predicted. For differences in curvature between the two conditions, $F(2, 414) = 12.95$, $p < .001$.

Standard deviation of bias. Recall that responses for the normal condition should be less variable than responses for the uniform condition, and responses for the narrow half uniform distributions of stimuli should be less variable than responses for the wide uniform condition. These predictions were tested in the same manner as in Experiments 1 and 2. The logs of the standard deviations in the normal and uniform conditions were compared for the 10 most central stimuli. The average log-transformed standard deviation in the normal condition was 3.926, and in the uniform condition it was 4.083. The difference is significant, $t(18) = 2.306$, $p < .05$. The logged standard deviations for the short and long half distributions were compared with standard deviations for comparable stimuli in the wide uniform condition. The average logged standard deviation in the short half distribution was 3.624, and for the corresponding portion of the wide uniform distribution it was 3.618. The difference was nonsignificant, $t(22) = -0.051$, $p > .95$; the result was similar when only the early trials of the short distribution were considered, mean logged standard deviation = 3.552, $t(22) = -0.446$, $p > .65$. The average transformed standard deviation in the long half distribution was 3.948, and the average for the corresponding stimuli in the wide uniform condition was 4.345. The difference was highly significant, $t(22) = 7.194$, $p < .00001$. The result was similar when only the early dark trials were considered: The mean for the dark half was 3.918, and the difference was highly significant, $t(22) = 7.807$, $p < .00001$. Thus, the variability of responses was lower for two out of three conditions in which information about the category was more precise. Table 3 summarizes the results.

Discussion

As in the previous experiments, responses were biased toward the center of the distribution, and increasing the precision of the central value by concentrating instances in the category led to less response variability. The dispersion of instances within a category affected the shape of the bias pattern. There was less bias for extreme stimulus values in the normal condition than in the uniform condition. Note, moreover, that the pattern of bias is consistent with the idea that length is a dimension with increasing magnitude, which would traditionally be represented using a log scale. As we discussed earlier, when stimulus uncertainty increases with stimulus magnitude, bias should take the form of an accelerating curve, as in Figures 8A and 8B. The downward concavity in 8A, and in 8B up to the point of the hook that begins at 315 pixels, seems more pronounced than in the previous experiments. Our predictions about the effect of stimulus condition on response variability were only partially supported in this experiment. Decreasing the range of presented instances (in the short half and long half conditions) resulted in the predicted reduction in variability only for the long half.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Variability in Experiment 3

Condition	*SD* in Pixels	Log *SD*	*t*	*df*	*p*
Normal	50.915	3.926			
Uniform	60.270	4.083			
Difference	9.355	0.157	2.306	18	.0332
Uniform	39.224	3.618			
Short half	38.699	3.624			
Difference	0.525	−0.006	−0.051	18	.9598
Uniform	77.664	4.345			
Long half	52.309	3.948			
Difference	25.355	0.397	7.194	22	3.283×10^{-7}

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

General Discussion

In this article we have evaluated a model that posits that category effects on stimulus judgment arise in pursuit of an adaptive goal—to maximize accuracy. According to the model, stimuli are encoded hierarchically, at fine-grain and category levels, and information at these two levels is combined in estimation. The notion that stimulus encoding is hierarchical and that people use prior information in interpreting situations is familiar. It is widely recognized that categories provide information about stimulus features that are not observed in particular situations. However, the use of category information to improve accuracy in estimating categorized stimuli has not been explored. Our model posits that because fine-grain stimulus representation generally is inexact, prior (category) information can be used to reduce the variability of estimates enough that accuracy is increased, even though some bias is introduced. People are unaware of adjusting stimulus values in forming estimates; the use of distributional information is automatic and ubiquitous.

Our earlier studies tested one prediction of the model—that the weight given to category information depends on the inexactness of stimulus representation. If representation is exact at the time of estimation, category information will not be used; if it is very inexact, category information will be important. This article verified another prediction of the model—that the effect of inductive categories on estimation depends on the nature of the observed distribution on which they are based. First, when the relative density of instances near category boundaries is low, as in a normal distribution, the probability of membership drops off more away from the category center than for a uniform distribution; hence the bias curve should be less linear, leveling off or decreasing at particular extreme locations. Second, when the concentration of instances in a category is greater, estimates of stimulus values should be less variable. The present results, together with our earlier results, strongly support the claim that the processes used in estimating inexactly represented stimuli are adaptive in that they achieve high average accuracy.

This article used single dimensional stimuli to test the model's predictions. Clearly, it is important to know whether this prediction about instance dispersion also holds for multidimensional stimuli. In a recent study, we used two-dimensional stimuli to test whether the concentration of instances in a category affects the variability of estimates (Crawford, Huttenlocher & Hedges, 2000). The study was based on earlier work by Huttenlocher and Hedges (1994) concerning the shapes of two-dimensional categories based on observed densities of stimulus values. Category shape, they noted, should reflect both the distribution of values on each of the dimensions and the correlation between the dimensions. When the dimensions are correlated, the category should be elliptically shaped, with axes that vary in width, as in Figure 9. In our study, stimuli varied along two dimensions (shade of gray and fatness of fish) that were correlated at .7. As predicted, an elliptically shaped category was formed that reflected the correlated distribution that had been presented; the variability of estimates measured along the minor axis of the ellipse was smaller than the variability measured along the major axis. This finding provides powerful additional evidence for the model's prediction that

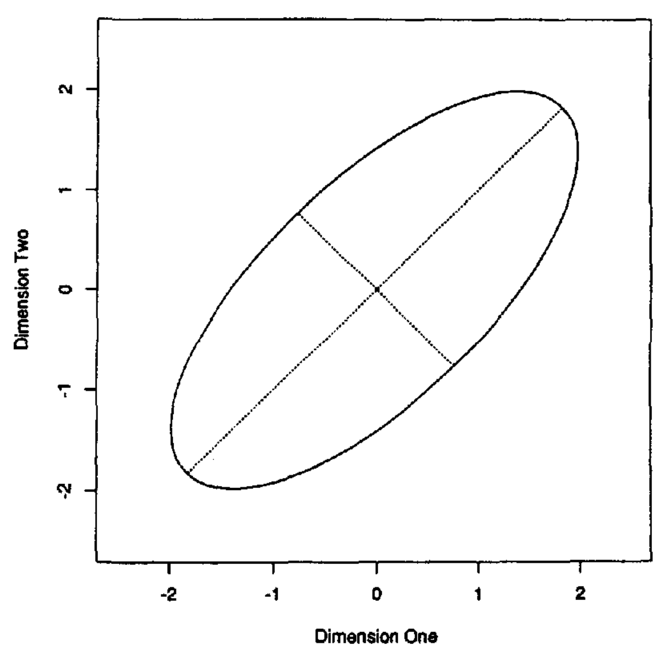

Figure 9. Schematic representation of a two-dimensional category formed from correlated dimensions.

the variability of estimates should be less when the concentration of instances is greater.

Alternative Methods for Studying Stimulus Judgment

The judgment of stimulus magnitudes has long been a focus of study by psychophysicists. The method of stimulus reproduction, used in our studies, has been used only rarely in psychophysical studies (excepting duration judgments). Because these studies, like ours, attempt to relate "sensation" (i.e., the mental impression of a stimulus) to physical stimulus values, it is important that the different methods yield consistent results. Stimulus reproduction provides a straightforward measure of stimulus judgment. The reason it has not commonly been used may be that the stimuli traditionally studied by psychophysicists cannot easily be reproduced (e.g., the weights of objects). Even for the visual stimuli we used, reproduction involved computers (e.g., reproducing shades of gray), but many psychophysical studies were done before computers were used in experimental work.

Numerical Scales

The most common psychophysical tasks involve the use of numerical scales. For example, people may be asked to use the numbers 1 through 5, 1 through 9, and so forth, to indicate line length from very short to very long. In some experiments each stimulus is associated with a different and unique number (called absolute identification), and in some cases there are fewer numbers than distinct stimuli (called category rating). The process of mapping mental impressions of stimuli onto numerical responses may introduce complexity beyond that involved in stimulus reproduction, so these tasks may provide a less straightforward measure of people's impression of a stimulus. Nevertheless, many of the effects found in psychophysics parallel our effects. Hollingworth (1910) proposed a "central tendency of judgment" that parallels the well-known schema or assimilation effect we are concerned with. Many psychophysical studies focus on contrast, which is generally found when stimuli in a set are judged relative to an anchor stimulus falling outside that set. The findings are consistent with our model; the stimuli within the set are adjusted toward the center of that set, whereas the anchor stimulus from outside the set is not adjusted. Therefore, anchor stimuli are judged as more different from members of the set than they truly are.

Not all findings of psychophysical studies are easily reconciled with our model. One such effect, explored by Parducci (1965), arises in tasks involving category rating. Judgment is affected by the response categories available. Parducci proposed a "range frequency principle" according to which people tend to equalize the use of available responses. Equalization of responses over an appropriate range does not predict bias for a uniform distribution and, for a normal distribution, response equalization would result in adjustment outwards, away from the center of the category, not inwards toward the center of the category, as we found. As Poulton (1989) suggested, this principle may apply only in experiments where rating scales are used and thus would not be expected in our reproduction task.

Discrimination

As we have seen, another method for evaluating category effects is to compare the discriminability of stimulus pairs that are equidistant in physical units when they are from a single category versus two adjoining categories. In our model, discrimination should be better for pairs that are from different categories rather than from the same category because comparisons are based on adjusted values. When stimuli are drawn from two categories along the common dimension, one may find two bias curves, one for each category, as shown in the schematic plot in Figure 10. Because the adjustment of stimuli in different categories is made toward different prototypes, a discontinuity (contrast) will be found across the category boundary. There will be underestimation on one side of the boundary and overestimation on the other side of the boundary, as we have demonstrated on stimulus reproduction tasks (e.g., Huttenlocher et al., 1991). Discriminability will be greater across a category boundary, because adjusted stimulus values are closer to category centers and farther from boundaries (and each other) than physical values. Engebretson (1995) showed this effect on a discrimination task using the same stimuli we used earlier on a reproduction task.

Alternative Explanations of Category Effects on Stimulus Estimation

Psychophysical studies generally have been concerned with showing mathematical relations of stimuli to judgment in various contexts, not with explanatory models. However, there are alternatives to our model of the sources of category effects. Let us consider two such alternative explanations and contrast them with our model.

Regression to the Mean

A statistical model has been proposed for category effects that, contrary to our model, holds that these effects can be

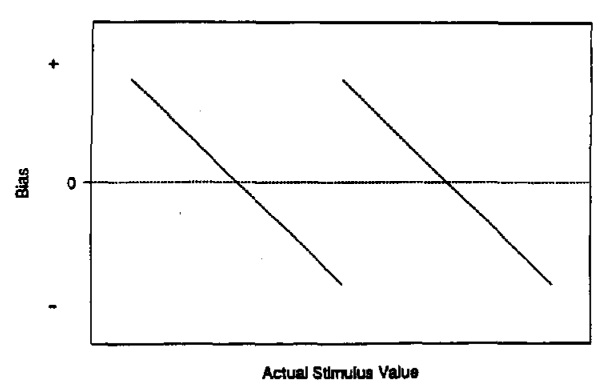

Figure 10. Schematic representation of bias due to prototype weighting for two adjacent categories.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

explained mathematically rather than requiring a psychological explanation at all. In this view, spelled out by Johnson and Mullally (1969), it is argued that the central tendency of judgment is due to *regression to the mean,* as described by Galton (1889). In Galton's formulation, regression to the mean arises because of measurement error, that is, because uncertain values for one measure on a scale (e.g., an individual's test score) are used to predict uncertain values for a second measure on that scale (e.g., a second score on that test). The statistical argument is that an extreme value on the first measure may be extreme not only because the *true* value is extreme but also because measurement error happens to fall in the same direction. Because an extreme measurement error is unlikely to repeat, the value on the second measure (the second test score) can be expected to be less extreme than that on the first.

Galton's principle does not apply to all cases of bias toward the mean. If the true stimulus value is the first measure (e.g., the actual loudness of a tone or size of a square, etc.) then it has no error. Therefore, error in this measure cannot be invoked to explain why values on the second measure are less extreme than those on the first.[4] Because this is the case in our experiments, another explanation of bias toward the mean is required. That is, the term *regression to the mean* can *describe* but does not *explain* the central tendency of judgment unless there is measurement error on both measures.

How Regression to the Mean Differs From Our Model

In our model too, statistical principles are invoked to explain the central tendency of judgment. However, the basis is different. Ours is a psychological explanation in which people use category information to maximize the accuracy of their estimates of inexactly represented stimuli. We examine whether people behave as Bayesians by determining whether their estimates are affected by the nature of the distributions in which stimuli are embedded. Our model, described mathematically in the Appendix, predicts the general form the data should take. In this way, our model also contrasts with many mathematical models in cognitive psychology, generally information processing models, where experimental data are used to estimate parameters necessary for more refined predictions about those data.

Learned Distinctiveness

Another proposed explanation of category effects, as we have noted, is that they are due to long-term changes in the representation of stimulus dimensions caused by learning a category. Most directly relevant here is Goldstone's (1994) argument that category learning leads to stretching of selected regions of a single dimension if that dimension is divided into more than one category. Following earlier work on phoneme discrimination, described as "categorical perception" of phonemes (e.g., Liberman et al., 1967), Goldstone obtained parallel effects for discrimination after teaching people categories, and he too interpreted the effects as perceptual.

Although Goldstone focuses on discrimination tasks, the notion that perception (encoding) is altered by category learning implies that only altered values are available to the cognitive system after learning. Although he does not consider estimation of the values of categorized stimuli, bias should be expected by his account because stimuli that were equidistant before learning become "stretched" in the learning process. A weakness of his proposal in explaining the results of tasks involving stimulus reconstruction is that it does not provide a mechanism by which variations in experimental conditions should affect stimulus judgment. Yet category effects on stimulus reconstruction increase when stimulus representation is less exact—when encoding conditions are degraded (Biederman, 1981), when stimuli are in memory longer (cf. Bartlett, 1932; Brewer & Nakamura, 1984), or when interference tasks are given (cf. Huttenlocher et al., 1991).

How Learned Distinctiveness Differs From Our Model

In our model, as in Goldstone's proposal, stimulus judgment is affected by category learning. However, the effect arises in making estimates rather than in the process of encoding. A category is a structure in memory; category information is used to improve the accuracy of judgment when fine-grain encoding is inexact. Because fine-grain and category information are both encoded, the category can be used immediately in estimation if the fine-grain value is inexact at that time. In addition, contrary to Goldstone's proposal, if estimates are made later, after a stimulus has been in memory for some time, they are likely to be more biased, because stimulus representation will have become less exact. In short, it is a tenet of our model that experimental conditions will affect the use of category information; the weight given to the category reflects stimulus inexactness at the time of estimation.

The present studies do not rule out the possibility that bias arises entirely during encoding. However, our earlier work provided evidence that stimulus adjustments are made at estimation; that is, bias increased when an interference task was introduced after the stimulus was presented and before it was reproduced (Huttenlocher et al., 1991). The categories in that study were not inductive categories. However, in recent work (Crawford, Engebretson, & Huttenlocher, 2000) we used stimuli that varied in length in a task where we expected to find two sorts of bias—one occurring at encoding and one at estimation. Lines of different lengths

[4] Galton introduced the concept of regression to the mean in connection with predictions of sons' heights from those of their fathers. He found that the sons' heights tended to be less extreme than their fathers' heights. The physical measurement of heights involves little error. Hence this might seem like our case, where statistical regression would not be expected. However, the underlying variable mediating the relation between fathers' and sons' heights is the genotype, whereas the measure is the phenotype; both variables are "measured with error" with respect to the genotype. It is error of measurement of the fathers' genotype (the first measure) that is the source of regression to the mean in Galton's example.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

were embedded in the two contexts of the Muller–Lyer illusion and in a neutral context. One group had lines that had arrows pointing in, another group had lines that had arrows pointing out, and a third group had no arrows. For each context, there were two different conditions—one where the line that was reproduced was present, and one where the line reproduced had been removed from view. In the line-present condition, bias due to Muller–Lyer arrows occurred for lines of all lengths, but there was no bias toward the mean length. In the line-absent condition, bias due to arrows was similar to that in the line-present condition, and there was bias toward the mean length in all three presentation contexts. Thus, in this experiment, bias toward a prototype did not arise at encoding like the Muller–Lyer effect, but only later.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Category Judgments

Our model concerns the use of relative frequency *within* a category. As we have noted, earlier investigators examined the use of relative frequency *across* categories in judgments of category membership. Our model, which posits that relative frequency information is preserved and used to improve accuracy, should affect both across-category judgments and within-category judgments. The existing literature suggests that use of across-category frequencies may be less ubiquitous than use of within-category frequencies. Note that our Bayesian model would predict just this, because rational category decisions are driven by their consequences, and these may differ across categories. Indeed, differences in consequences is a major reason for partitioning objects into different categories (i.e., people may form distinct categories of poisonous mushrooms and edible mushrooms).

When category errors have unequal costs, accuracy may not be the most important principle. Given the high cost of wrongly classifying poisonous mushrooms as edible, mushrooms may be treated as poisonous if there is any possibility that they belong in that category, or, as in the example in Koehler's (1996) article, people may be treated for a rare serious disease, even though benign conditions are more frequent, because the consequences of misdiagnosis are dire. Only when the consequences of mistakes are equally disadvantageous is relative frequency the critical principle in category decisions. Because within a category there is generally a commonality of consequences, accuracy is the important principle in stimulus estimation; relative frequency contributes to average accuracy so adjustment of stimuli is generally rational.

Further, cross-category frequencies are relevant to category decisions only when category membership is uncertain. When objects fall clearly within a particular category, as Koehler has noted, cross-category differences in frequency do not affect categorization. That is, even for infrequent categories, such as rare diseases like hemophilia or rare animals like aardvarks, unambiguous stimuli will be reliably categorized. In contrast, within-category frequencies can be used to improve the accuracy of estimation for all the stimuli. This is especially true when a stimulus is clearly within a category, but it is even true when there is uncertainty regarding category membership.

Conclusions

Our model, which posits that bias results from an adaptive mechanism that serves to increase the accuracy of judgment, led us to do a set of studies where people reproduce a series of stimuli. The model holds that people use prior (category) information to improve the accuracy of their estimates of inexactly represented stimuli. It makes predictions as to how variation in the distribution of stimuli in an inductive category should affect stimulus estimation. We verified these predictions in this article. Our model makes only skeletal assumptions about representation and processing because our first step has been to find evidence for the claim that the cognitive system operates so as to increase the average accuracy of estimates of categorized stimuli. Having found such evidence, our next step will be to develop more detailed information-processing models to further specify the mechanisms by which these adaptive ends are achieved.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

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

Ashby, F. G., & Lee, W. L. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General, 120,* 150–172.

Ashby, F. G., & Lee, W. W. (1992). On the relationship among identification, similarity, and categorization: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General, 121,* 385–393.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11,* 211–227.

Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition: Vol. 3: Content and process specificity in the effects of prior experiences* (pp. 61–88). Hillsdale, NJ: Erlbaum.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

Bartlett, M. S., & Kendall, D. G. (1946). The statistical analysis of variance-heterogeneity and the logarithmic transformation. *Journal of the Royal Statistical Society, 8* (Suppl.), 128–138.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization.* Hillsdale, NJ: Erlbaum.

Brewer, W. F., & Nakamura, G. V. (1984). The nature and function of schemas. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 119–160). Hillsdale, NJ: Erlbaum.

Crawford, L. E., Engebretson, P. H., & Huttenlocher, J. (in press). Perceptual and category bias in reproducing visual stimuli. *Psychological Science.*

Crawford, L. E., Huttenlocher, J., & Hedges, L. V. (2000). *Correlated attributes and category structure: The use of correlation in stimulus estimation.* Unpublished manuscript.

Engebretson, P. H. (1995). *Category effects on the discriminability of spatial location.* Unpublished doctoral dissertation, University of Chicago.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18,* 500–549.
Flanagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 241–256.
Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257.
Galton, F. (1889). *Natural inheritance.* London: Macmillan.
Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 23,* 178–200.
Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–447). New York: Bobbs Merrill.
Hollingworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology, and Scientific Methods, 7,* 461–469.
Homa, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 50–94). San Diego, CA: Academic Press.
Huttenlocher, J., & Hedges, L. V. (1994). Combining graded categories: Membership and typicality. *Psychological Review, 101,* 157–165.
Huttenlocher, J., Hedges, L. V., & Bradburn, N. M. (1990). Reports of elapsed time: Bounding and rounding processes in estimation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 196–213.
Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352–376.
Huttenlocher, J., Hedges, L. V., & Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review, 95,* 471–484.
Johnson, D., & Mullally, C. R. (1969). Correlation-and-regression model for category judgments. *Psychological Review, 76,* 205–215.
Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454.
Kalish, M. L., & Kruschke, J. K. (1997). Decision boundaries in one-dimensional categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1362–1377.
Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language, 54,* 610–646.
Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19,* 1–17.
Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 3–26.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.
Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics, 53,* 49–70.
Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.
Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27,* 148–193.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.
Nosofsky, R. M. (1998). Optimal performance and exemplar models of classification. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition.* New York: Oxford University Press.
Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381–391.
Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407–418.
Poulton, E. C. (1989). *Bias in quantifying judgments.* Hillsdale, NJ: Erlbaum.
Quinton, A. (1957). Properties and classes. *Proceedings of the Aristotelian Society* (Suppl.), 33–58.
Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7,* 532–547.
Tajfel, H. (1959). Quantitative judgment in social perception. *British Journal of Psychology, 54,* 101–114.
Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 60,* 471–480.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix

Mathematical Model Created by Larry V. Hedges and Jack L. Vevea

First we present the notion of an inexactly represented fine-grain stimulus value. Then we present the uses of prior information as posited in the model. We start with the case where the area of inexactness of a fine-grain value falls entirely within the boundaries of a category. We show how combining an inexact memory with the central value of a category (a prototype) affects the bias and standard deviation of responses, potentially increasing accuracy. Then we consider the predictions tested in this article. We also show how variation in a prior distribution should affect estimation. We show how variation in the concentration of instances in a category affects the bias and standard deviation of responses. We consider variation in dispersion and how the adjustment due to category effects is affected when the boundaries of a category are uncertain.

I. Inexactness of Fine-Grained Memories

Our model distinguishes responses (*R*s) from the fine-grained memory (*M*). The idea of inexactness in *M* is operationalized by treating it as a random variable with standard deviation σ_M. We posit that *M* has a normal distribution with mean μ, which for representations of a particular stimulus is equal to the true value for that stimulus. An alternative description is that the representation for a particular stimulus is $M = \mu + \epsilon$ where ϵ is a random deviation with variance σ_M^2.

II. Effects of Adjustment Toward Central Values of a Category

Our model posits that the *R* is a linear combination of the uncertain *M* (as in Section I) and the central value of the category (ρ). The inexactness of the central value is operationalized as the standard deviation of a random variable that reflects the dispersion of values in the category. Note however that this inexactness (denoted σ_P) of the central value is not the uncertainty of the value ρ, which may be known with great precision, but the inexactness of ρ as an estimate of an arbitrarily chosen category value. In Bayesian terms, σ_P reflects the standard deviation of the prior distribution, not the standard error of the prior mean.

The weight (λ) given to *M* is a smooth (differentiable) monotonic function *g* of the ratio of memory uncertainty (σ_M) to category concentration (σ_P). Thus, the response is

$$R = \lambda M + (1 - \lambda)\rho,$$

where $\lambda = g(\sigma_M/\sigma_P)$ is a monotonic decreasing function whose range is from 0 to 1. We assume that when the memory uncertainty is zero, $\lambda = 1$, so that the response is entirely determined by *M*. Similarly, we assume that when the relative uncertainty of memory is sufficiently large, the response is determined entirely by the central category value. Thus $g(0) = 1$ and there exists a (large) value *c*, such that $g(c) = 0$. These considerations are sufficient to demonstrate the properties of the bias and standard deviation of *R* that are used in this paper.

The bias of *R* is simply the expected value of *R* minus the actual stimulus value, or

$$\mathrm{B}(R) = \mathrm{E}(R) - \mu = (\lambda - 1)(\mu - \rho),$$

where μ is the actual stimulus value. Thus, although *M* is unbiased, *R* is biased toward the central value of the category. The standard deviation of *R* is simply

$$\mathrm{S}(R) = \lambda\sigma_M,$$

which implies that $\mathrm{S}(R) \leq \sigma_M$ and adjustment reduces variability because $\lambda \leq 1$. Below we show that *R* may be more accurate than *M* as an estimate of μ, and we derive the properties of the variance and bias of *R* as σ_P is held fixed and σ_M is varied and (in Section VI) as σ_M is held fixed and σ_P is varied.

III. Category Effects When Memory Inexactness (σ_M) Is Varied

In Section II we showed that using a weighted combination of fine-grain and category information could increase accuracy. In this section we determine the effects on the components of accuracy (bias and variance) of varying memory uncertainty.

A. Effects on bias. To obtain the properties of the bias, we compute the partial derivative of B(*R*) with respect to σ_M, which yields

$$\mathrm{B}'(R) = g'(\sigma_M/\sigma_P)(\mu - \rho).$$

Because $g(x)$ is a strictly decreasing function, $g'(x) < 0$ and therefore $\mathrm{B}'(R)$ is positive when $\mu < \rho$ and negative when $\mu > \rho$. Thus B(*R*) is strictly increasing whenever $\mu < \rho$ (where bias is positive) and strictly decreasing whenever $\mu > \rho$ (where bias is negative), so that increasing σ_M always increases the absolute magnitude of the bias. Figure 1A shows the effect on the bias of responses (for a fixed stimulus below the category center) of changing σ_M from 0 to a very large value with σ_P fixed. As expected, the bias increases monotonically from 0 for small values of σ_M to $\mu - \rho$ for large values of σ_M.

B. Effects on variability. To obtain the properties of the standard deviation of the response, we note that the partial derivative of S(*R*) with respect to σ_M,

$$\mathrm{S}'(R) = \lambda + g'(\sigma_M/\sigma_P)\sigma_M,$$

need not be positive, so that S(*R*) need not be monotonic. In fact, S(*R*) is never monotonic as a function of σ_M because S(*R*) always has a maximum. Recall that $g(0) = 0$, so $\mathrm{S}(R) = 0$ when $\sigma_M = 0$. Recall also that there is a large value *c* such that for $(\sigma_M/\sigma_P) = c$, $g(c) = 0$, which implies $\mathrm{S}(R) = 0$ for $\sigma_M = c\sigma_P$. The intermediate value theorem implies that there exists a point *x* between 0 and $c\sigma_P$ such that $g'(x) = 0$. Thus there is a point *x* where S(*R*)

(*Appendix continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

has a maximum, so $S(R)$ cannot be monotonic as a function of σ_M.

Figure 1B shows the effect on the standard deviation of responses when optimal weighting is used. The standard deviation of R exhibits the expected pattern of first increasing to a maximum and then declining for larger values of σ_M.

IV. Improving Accuracy

The accuracy of an estimate is the average or expected value of the squared difference between the estimate and the true value. The accuracy of an estimate can be shown to be the sum of the square of the bias and the variance. The results of Section II can be combined to give an expression for the accuracy (mean square error) of R in terms of λ, μ, ρ, and σ_M^2:

$$MSE(R) = (\lambda - 1)^2(\mu - \rho)^2 + \lambda^2\sigma_M^2.$$

Therefore, solving for the values of λ that give $MSE(R) < \sigma_M^2$ gives the values of λ for which the adjusted value R is more accurate than the unadjusted inexact value from memory. The expression for $MSE(R)$ implies that $MSE(R) < \sigma_M^2$ is equivalent to

$$(\lambda - 1)^2(\mu - \rho)^2 + (\lambda^2 - 1)\sigma_M^2 < 0.$$

Collecting terms and using the quadratic formula to solve the corresponding equation for λ yields that $MSE(R) < \sigma_M^2$ whenever

$$[(\mu - \rho)^2 - \sigma_M^2]/[(\mu - \rho)^2 + \sigma_M^2] < \lambda < 1.$$

Thus, there is always a range of adjustment weights λ (including the optimal adjustment) that produce an adjusted estimate that is more accurate than the unadjusted, but unbiased, inexact value from memory. The optimal adjustment can be derived from Bayes's theorem. When the distribution of instances is normal (or uniform), the optimal adjustment weight for the uncertain memory is proportional to the reciprocal of the memory uncertainty or

$$\lambda = \sigma_P^2/(\sigma_M^2 + \sigma_P^2).$$

V. Pattern of Bias Across the Category

In Section II we showed that the bias is

$$B(R) = (\lambda - 1)\mu + (1 - \lambda)\rho,$$

and argued that λ was a monotonic decreasing function of the ratio of memory uncertainty σ_M to concentration σ_P. If σ_M is constant across the category, λ will therefore be constant across the category and bias will be a linear function of stimulus location μ. Because $\lambda < 1$, $\lambda - 1$ is negative, and the bias as a function of μ will have negative slope as illustrated in Figure 2A. Alternatively, if memory uncertainty increases with stimulus size, then λ will decrease with stimulus size and bias will decrease at a rate that increases as μ increases, as illustrated in Figure 2B. Note that any changes that affect the category as a whole (such as changes in experimental conditions) may change the slope or rate of change of the bias, but they will not affect the general shape (e.g., linear or not) of the bias as a function of stimulus size.

VI. Category Effects When Membership Is Uncertain

A. Response bias. Uncertainty about category membership affects shrinkage toward the central value of the category for stimuli that are sufficiently far from the center of the category. To make explicit the effects of uncertain category membership, we must first describe the model that applies when category membership (or not) is certain.

Our model posits that when it is certain that a stimulus is a member of a category, a response is generated by the process of mixing fine-grain memory with category information, as we described. When it is certain that a stimulus is *not* a member of the category, the response is based entirely on the fine-grain memory. Thus, the bias of a response is simply

$$B(R) = (\lambda - 1)(\mu - \rho)$$

if the stimulus is a category memory, or 0 if not.

The standard deviation of responses is

$$S(R) = \lambda\sigma_M$$

if the stimulus is a category member, or σ_M if not.

When it is uncertain whether a stimulus is a category member, there exists some probability π (a function of μ) that the stimulus with true value μ is considered to be a category member. As before, to the extent that a stimulus is a member of the category, the fine-grain memory will be adjusted using category information; to the extent that the stimulus is considered a nonmember of the category, the response will be the fine-grain memory. Therefore, if category membership is uncertain, the response is

$$R = (1 - \pi)M + \pi[\lambda M + (1 - \lambda)\rho]$$

and the bias of the responses then becomes

$$B(R) = \pi(1 - \lambda)(\rho - \mu),$$

and the standard deviation of the responses is

$$S(R) = [1 - \pi(1 - \lambda)]\sigma_M.$$

B. Improving accuracy when category membership is uncertain. To show that this adjustment is rational consider the mean squared error of R conditionally (on the decision about category membership) where the random variable $C = 1$ if a response is derived from an inexact value (M) classified as certainly being in the category and $C = 0$ if not. Thus the mean squared error of R is

$$MSE(R) = MSE(R|C = 1)P(C = 1) + MSE(R|C = 0)P(C = 0),$$

and the mean square error of M is

$$MSE(M) = MSE(M|C = 1)P(C = 1) + MSE(M|C = 0)P(C = 0),$$

where the notation $P(A)$ is the probability of event A, and $MSE(R|A)$ is the conditional mean squared error given A. Because the second terms of the righthand sides of the expression are identical, but $MSE(R|C = 1) < MSE(M)$, as shown in Section IV, $MSE(R) < MSE(M)$, this adjustment is more accurate than use of the unadjusted inexact memories. Considering the probability π as a prior distribution of category membership and applying Bayes's

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

theorem leads to the conclusion that if λ is the optimal shrinkage factor given in Section IV, then R is a Bayes estimate and therefore optimal. Because the probability of category membership decreases with stimulus inexactness for stimuli away from the center of the category, the expression for the bias given in this section implies that inexactness will decrease bias by decreasing π, the probability of membership.

VII. Effects of Category Adjustment When the Dispersion of Instances (σ_P) Is Varied

In Section IV we showed that a weighted combination of fine-grain memory and the central value of a category could increase accuracy. In this section, we determine the effects on the components of accuracy (bias and variance) of varying the distribution of instances in the category (σ_P).

A. Effects on variability. To obtain the properties of S(R), we compute the partial derivative of S(R) with respect to σ_P

$$S'(R) = -(\sigma_M/\sigma_P^2)g'(\sigma_M/\sigma_P).$$

Because $g(x)$ is a strictly decreasing function, $g' < 0$ and therefore S′(R) is always positive; hence increasing σ_P always increases S(R).

Figure 3A shows the effect on standard deviation of R of changing σ_P from 0 to a very large value (with σ_M fixed). One complication in producing plots where σ_P is varied is that one cannot fix the absolute position of the stimulus as σ_P tends to zero. The reason is that if σ_P is sufficiently small any fixed stimulus value will be too far from the prototype to be possible given the small size of σ_P. Consequently, the standard deviations of responses are plotted for a stimulus at the same relative position in the category (one standard deviation below the category prototype) as σ_P increases. The standard deviation of R increases monotonically, approaching the value σ_M as σ_P becomes very large.

B. Effects on bias. Because σ_P depends on μ, the partial derivative of B(R) with respect to σ_P is difficult to compute. However, B(R) is never monotonic as σ_P is varied because B(R) always has a maximum. Note that when $\sigma_P = 0$, the category has collapsed to a single value and hence $\mu = \rho$, so B(R) = 0 when $\sigma_P = 0$. Recall that when there is a large value c such that for $(\sigma_M/\sigma_P) = c$, $\lambda = g(c) = 0$, which implies that B(R) = 0 for $\sigma_P = \sigma_M/c$. The intermediate value theorem implies that there exists a point x between 0 and σ_M/c such that $g'(x) = 0$. Thus there is a point where B(R) has a maximum so that B(R) cannot be monotonic as a function of σ_P. Figure 3B depicts the relation when optimal weighting is used: The bias of R first increases for small values of σ_P (plotted as the one-fourth power of σ_P to better illustrate the pattern), reaches a maximum, and then decreases for larger values of σ_P.

Received June 18, 1998
Revision received February 24, 1999
Accepted July 14, 1999 ■

New Editors Appointed: *Emotion*

The Publications and Communications Board of the American Psychological Association announces the appointment of Richard J. Davidson, PhD (Department of Psychology, University of Wisconsin—Madison), and Klaus R. Scherer, PhD (Department of Psychology, University of Geneva), as co-editors for the new APA journal *Emotion* for the term 2001–2006.

Effective immediately, please submit manuscripts (five copies) to

Richard J. Davidson, PhD
Emotion Journal Office
Department of Psychology and Waisman Center
University of Wisconsin—Madison
1500 Highland Avenue
Madison, WI 53705-2280