Journal of Experimental Psychology: General
2007, Vol. 136, No. 3, 502–519

0096-3445/07/$12.00 DOI: 10.1037/0096-3445.136.3.502

Estimating Stimuli From Contrasting Categories: Truncation Due to Boundaries

Janellen Huttenlocher, Larry V. Hedges, and Stella F. Lourenco
University of Chicago

L. Elizabeth Crawford
University of Richmond

Bryce Corrigan
University of Michigan

In this article, the authors present and test a formal model that holds that people use information about category boundaries in estimating inexactly represented stimuli. Boundaries restrict stimuli that are category members to fall within a particular range. This model posits that people increase the average accuracy of stimulus estimates by integrating fine-grain values with boundary information, eliminating extreme responses. The authors present 4 experiments in which people estimated sizes of squares from 2 adjacent or partially overlapping stimulus sets. When stimuli from the 2 sets were paired in presentation, people formed relative size categories, truncating their estimates at the boundaries of these categories. Truncation at the boundary of separation between the categories led to exaggeration of differences between stimuli that cross categories. Yet truncated values are shown to be more accurate on average than unadjusted values.

Keywords: Bayesian, boundaries, categories, estimation, truncation

Categorization is enormously important in cognitive functioning. Because “objects or events are alike in some respects,” it is possible for particular experiences to “be thought about and responded to in ways . . . already mastered” (Smith & Medin, 1981, p. 8). Categories preserve critical information about the range and distribution of instances (extensional information) and nonobvious “hidden” attributes of instances (intensional information). However, they are also a source of bias in stimulus judgment. For example, categorized stimuli are often judged as being closer to a category center than they truly are, and a pair of stimuli from a common category may be judged as being more similar than an equidistant pair from different categories. There is an extensive literature on bias stemming from categories in many domains—object categories, social categories, and sets (categories) of experimental stimuli among others (e.g., Bartlett, 1932; Brewer & Nakamura, 1984; Hollingworth, 1910; Poulton, 1989; Tajfel, 1959). These two notions about categories—that they provide critical information about stimuli, yet lead to bias in judging those stimuli—seem to be incompatible.

Janellen Huttenlocher and Stella F. Lourenco, Department of Psychology, University of Chicago; Larry V. Hedges, Department of Sociology, University of Chicago; L. Elizabeth Crawford, Department of Psychology, University of Richmond; Bryce Corrigan, Department of Political Science, University of Michigan.

Larry V. Hedges is now at the Department of Statistics and the Department of Education and Social Policy, Northwestern University.

This research was supported in part by a grant from the National Science Foundation (REC-0087516). We thank Yaakov Kareev, Terry Regier, and Douglas Wedell for their comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Janellen Huttenlocher, Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637. E-mail: hutt@uchicago.edu

Our research suggests a resolution of this seeming contradiction. Category bias, we argue, reflects a process in which inexact information at two levels of detail (fine-grain and category levels) is combined. Combining these partially redundant sources of information introduces bias in estimation but can increase the accuracy and stability of stimulus estimates (e.g., Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000). The extent to which category information can increase accuracy of estimates is greater when fine-grain values are less exact. One category adjustment process is weighting with a prototype, in which information about a central value is integrated with fine-grain values. This process has a direct parallel in a well-known procedure in Bayesian statistics: Stein shrinkage (e.g., Huttenlocher et al., 2000).

The present article concerns *truncation*, a category adjustment process in which information about category boundaries is integrated with fine-grain values. This process eliminates stimuli recalled as falling outside the category. Although truncation is a well-known phenomenon, its implications for accuracy of estimation have not previously been addressed. Here we show that truncation can be treated as a Bayesian procedure that increases the average accuracy of estimates. In the Appendix of this article, we outline the truncation model (Section 1), indicating how the parameters are fitted from the data (Section 2). Then we present arguments to show that truncation involves optimal use of prior information about category boundaries (Section 3) and that it improves accuracy even in suboptimal conditions (Section 4).

In the “classic” view, categories are defined by boundaries that specify the necessary and sufficient (defining) conditions for membership. All stimuli falling within the boundaries are equally good members (cf. Smith & Medin, 1981). It is widely recognized, however, that there are difficulties with this view, because many categories lack clear-cut boundaries, and instances are better or worse depending on their relation to a prototype or "best" instance (e.g., Posner & Keele, 1968). Even when boundaries are uncertain and category decisions are probabilistic, though, categories are projectible (Quinton, 1957). They provide the basis for decisions about the membership of stimuli that are important to the accuracy of stimulus estimation. That is, a correct decision leads to adjustment of stimuli toward the center of the proper category, whereas a mistaken decision leads to adjustment of stimuli in the wrong direction, with potentially large negative effects on accuracy.

In earlier work, we showed truncation in people's reports of the dates of movies that they saw at the University of Chicago over a school year (Huttenlocher, Hedges, & Prohaska, 1988). People typically knew the academic quarter when movies were shown and the beginning and ending dates of the quarters. However, their memories for the particular dates of movies were quite inexact. We found a net forward bias in estimation; movie dates were reported as nearer the present than they actually were, a phenomenon referred to as *forward telescoping* in the survey literature (cf. Sudman & Bradburn, 1973). One could explain the observed bias by positing inexact but unbiased memories for the dates of movies, with truncation at the boundaries of academic quarters. Because memories become increasingly inexact over time, we argued that forward bias from an earlier boundary should be larger than backward bias from a later boundary. Although this work showed that category boundaries could be a source of bias, we had not yet developed the notion that truncation at category boundaries can increase the average accuracy of estimates.

In the present article, we apply the truncation model to the estimation of stimuli from contrasting pairs of categories, showing that even though truncation leads to bias, it increases average accuracy. We also show how truncation can explain effects that arise when semantic domains are partitioned into distinct categories. Such categories are distinguished by cues such as names or other associated characteristics. One well-known type of partitioning involves division into contrasting pairs of categories (e.g., for animals, warm-blooded vs. cold-blooded; for mushrooms, edible vs. poisonous). For such contrasting pairs of categories, the differences between stimuli that cross category boundaries may be exaggerated (e.g., Goldstone, 1996; Tajfel & Wilkes, 1963). We show how such exaggerations (bias) can result from a truncation process that increases accuracy of stimulus estimates.

THE CATEGORY ADJUSTMENT MODEL: TRUNCATION PROCESSES

We propose a category adjustment model that holds that stimulus representation is multilevel, involving partially redundant information at different levels. The general category adjustment model holds that integrating a sampled fine-grain value with higher level category information involves Bayesian statistical procedures that increase the average accuracy of estimates. One of these procedures is truncation. Below we briefly describe the assumptions of the model about the representation of fine-grain and category information and lay out the presumed mental processes involved in truncation; we refer to these as the *truncation model*.

Representation

According to the model, fine-grain stimulus information is unbiased but inexact. A stimulus consists of a value on dimensions such as the height of an object. The represented value is centered at the true value of the stimulus; inexactness around that stimulus forms a normal distribution. Such inexactness varies with stimulus magnitude (e.g., width, length, loudness). Representation becomes proportionally less exact as magnitude increases, as described by the Weber fraction. The relation of stimulus inexactness to magnitude can be captured by two parameters: an intercept and a slope that indicates how inexactness increases with magnitude. Dimensions of magnitude are asymmetric; they generally are bounded at the lower end by zero but are unbounded at the upper end. As Clark (1973) noted, this asymmetry is captured in linguistic coding. When we ask, "How long/wide/loud is it?" people answer using dimensions of length, width, or loudness, not of shortness, narrowness, or softness. This asymmetry will be relevant in our discussion of categories involving dimensions of magnitude.

A category consists of a region of a stimulus space with a particular distribution of instances and can be described by summary statistics that specify boundaries, central values, and dispersion of instances (Anderson, 1990; Ashby & Lee, 1991; Fried & Holyoak, 1984; Homa, 1984). When stimuli cluster in a stimulus space with few values in surrounding regions, the observed distribution may be captured in a category. Boundaries are estimates of the extreme values of the distribution. For unidimensional categories, there may be two boundaries (lower and upper). Because they are estimated, boundaries are imprecise to some degree—more so for larger than for smaller values.

Regardless of how instances are distributed over a stimulus space (domain), that space may be subdivided into distinct regions by category cues. That is, even if stimuli do not form distinct clusters, boundaries of separation may be formed where category cues shift. Category cues may be words (names) or other characteristics common to category members. For example, the differentiation of peaches from nectarines may have been formed via association with different names. The ways names are paired with instances establishes category extension—the range and distribution of instances—and differential use of names can carve out categories differently. Here we examine estimation of stimuli that are subdivided into two sets by category cues that cover distinct ranges of size.

Estimation

The category adjustment model holds that, in estimation, a fine-grain value is retrieved from the distribution of inexactness around the true value. In addition, category information is retrieved. There are two processes by which categories are used to adjust an inexact fine-grain value. One is weighting an inexact value with the mean of earlier instances, a process we have examined in earlier articles (Huttenlocher et al., 1991, 2000). The other is truncation in which values that fall outside the range of the category are eliminated, a process studied in the present article.

Truncation

When a true stimulus value is near a category boundary, a person may sample a fine-grain value that falls outside the range of the category. This value may be rejected and another sample drawn, because the original sampled value is probably wrong. Clearly, rejecting values that would place estimates outside the proper category improves accuracy. This intuition is clarified by an example. Consider a Mr. Brown who has met a basketball player named Joe. Mr. Brown is aware that basketball players tend to be tall. However, he met Joe long ago and remembers him only inexactly; in fact, he recalls Joe as relatively short. He may then reject this remembered value for Joe and sample again. If the inexactness of memory for Joe's height is centered at the true value, the distribution with resampling will be truncated, eliminating shorter values, and Joe will be estimated to be taller than he truly is. Nevertheless, on average, an individual's estimates of the heights of basketball players would be improved if he or she uses categorical information (assuming that the category accurately reflects observed stimuli).

Figure 1 shows truncation for stimuli at two locations near a lower boundary. Consider what happens to the distribution of stimulus uncertainty after truncation. Although the distribution is centered at the true value, the average of the reported values after truncation (based on values to the right of the boundary) is always larger than the true value because of rejection of values lower than the boundary. Bias is largest for stimuli nearest the boundary because the portion of the distribution that overlaps the boundary and thus is rejected is the greatest. The mathematical form of the distribution of estimates subject to truncation at a lower boundary is given in Section 1 of the Appendix. This mathematical form permits calculation of the bias introduced by truncation as a function of the stimulus inexactness and its location relative to the boundary.

Figures 2 and 3 are schematic graphs that show the results of truncation on response patterns. Figure 2 shows truncation at the lower boundaries of two adjacent categories. Unbiased values would lie at the zero point across the entire category. Bias from a lower boundary is always positive; smaller values are moved upwards because of truncation. It is also possible for there to be bias only at an upper boundary. In such a case, the bias would be negative and would be a function of the stimulus inexactness and location relative to the upper boundary, as indicated in Section 1 of the Appendix.

Figure 3 shows the impact of truncation at both upper and lower boundaries. The pattern of bias from an upper boundary is the opposite of that from a lower boundary: it is negative, and bias is greatest for the stimulus values near the upper boundary (large stimuli). Again, if reported values were unbiased, they would lie on a 45° line. Instead, with truncation at both lower and upper boundaries, average reported values for smaller actual stimulus values lay above the 45° line, and average reported values for larger actual stimulus values lay below it. The bottom graph shows bias on the vertical axis.

Truncation Under Varying Conditions

Additional processing factors arise for categories involving magnitude scales, and these have implications for bias due to

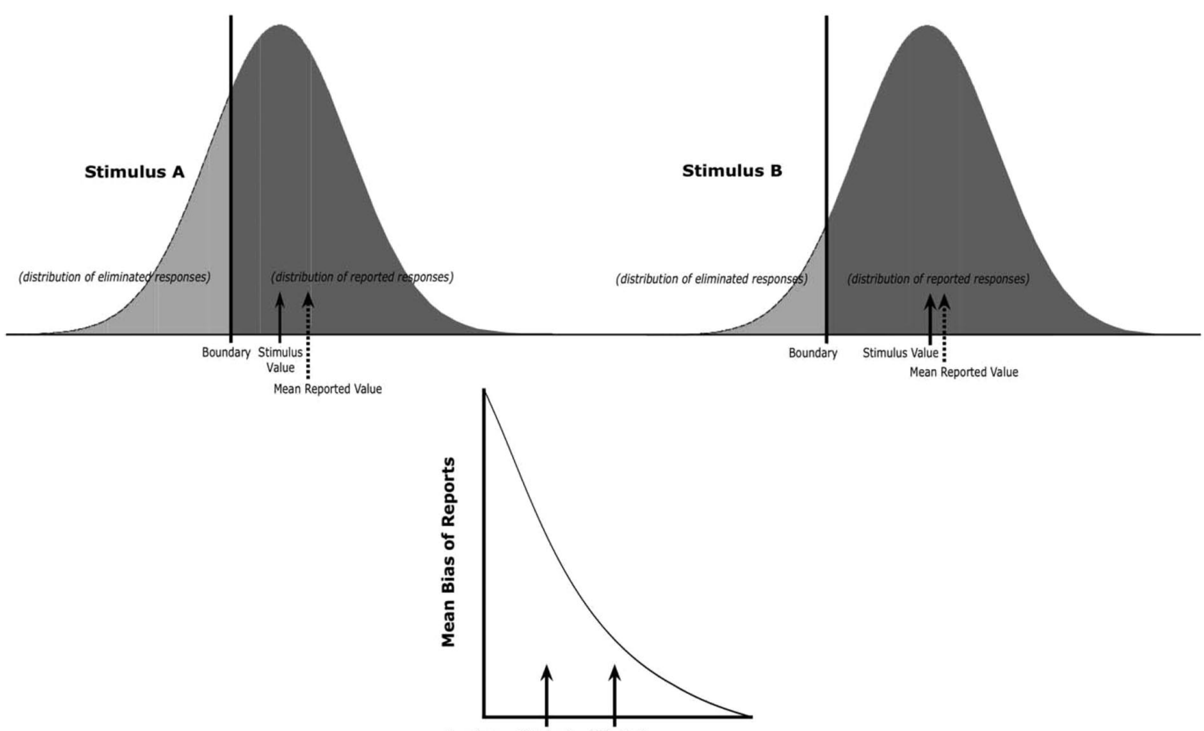

Figure 1. Graphs showing how truncation leads to bias: Effects at two different stimulus values and the functional relationship of stimulus location bias. The left panel shows a stimulus nearer the boundary (with greater overlap and larger bias in reported values), and the right panel shows a stimulus farther from the boundary (with less overlap and smaller bias in reported values). The bottom panel shows the functional relation between stimulus location and amount of bias in reported values.

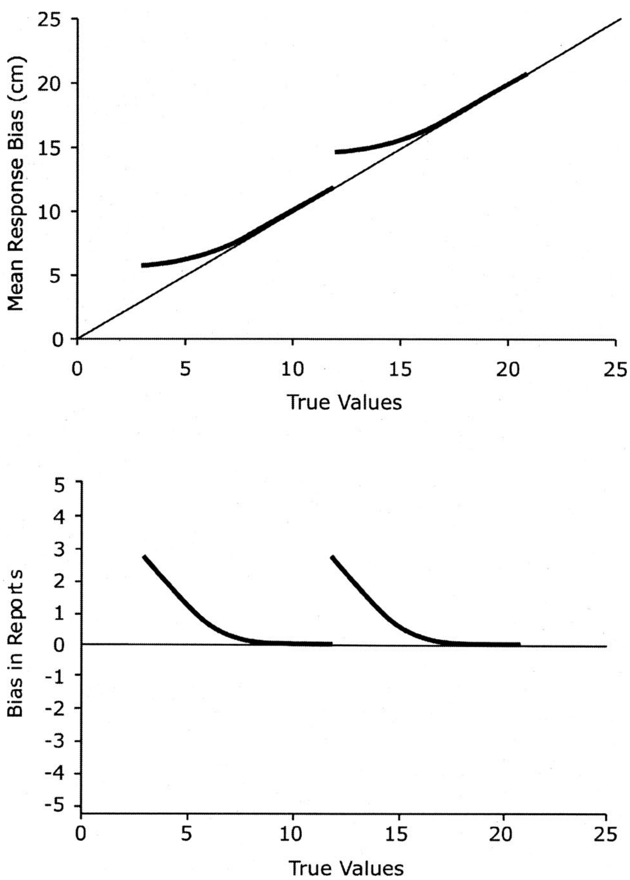

Figure 2. Graphs showing effect of truncation at the lower boundaries of two adjacent categories. The schematic graph in the top panel shows actual stimulus values on the horizontal axis and average reported values on the vertical axis. Unbiased reported values would lie on a 45° line. Instead, lower values are above the 45° line, indicating adjustment away from the lower boundary because of truncation. The graph in the bottom panel shows bias of reported values on the vertical axis (i.e., the true stimulus value minus the average observed value).

truncation. The extent of truncation from a boundary increases with the inexactness of stimulus representation at that location. Inexactness is related to stimulus size (as well as to coding and memory conditions). Given that representation is less exact for larger stimuli, truncation should be greater at larger values. Other things being equal, if two stimuli are in corresponding positions in categories that vary in magnitude, truncation would be greater for the stimulus in the larger category, and if two stimuli are in the same category, truncation would be greater for a stimulus nearer the upper rather than the lower boundary.

However, boundaries of inductive categories are themselves less exact for larger stimuli. The boundaries are constructed from the largest (or smallest) values in a set of encounters with category members. Because the representation of stimuli themselves is inexact (even more so for larger stimuli), boundaries based on larger stimuli are more inexact than those based on smaller stimuli. Truncation at inexact boundaries can be conceived of as the (weighted) average of truncation effects that would arise at each of a range of precise boundary values, in which the weights are determined by the likelihood of occurrence of that boundary value. Inexact boundaries cover a range of values, which leads to truncation effects that are less sharp than those for exact boundaries. If a boundary is very inexact, effects of truncation may not be observable.

Further, truncation at upper category boundaries may be less likely than at lower category boundaries because of the asymmetry of magnitude scales. As noted above, dimensions that form magnitude scales code the accumulation of extent at various points along a dimension. They are bounded at the lower end by zero; at the upper end, they are unbounded because extent can increase without limit. Hence, use of upper category boundaries in estimation may be less likely than use of lower boundaries.

The mathematical form of the distribution of estimates, subject to truncation at both a lower and an upper boundary, with different

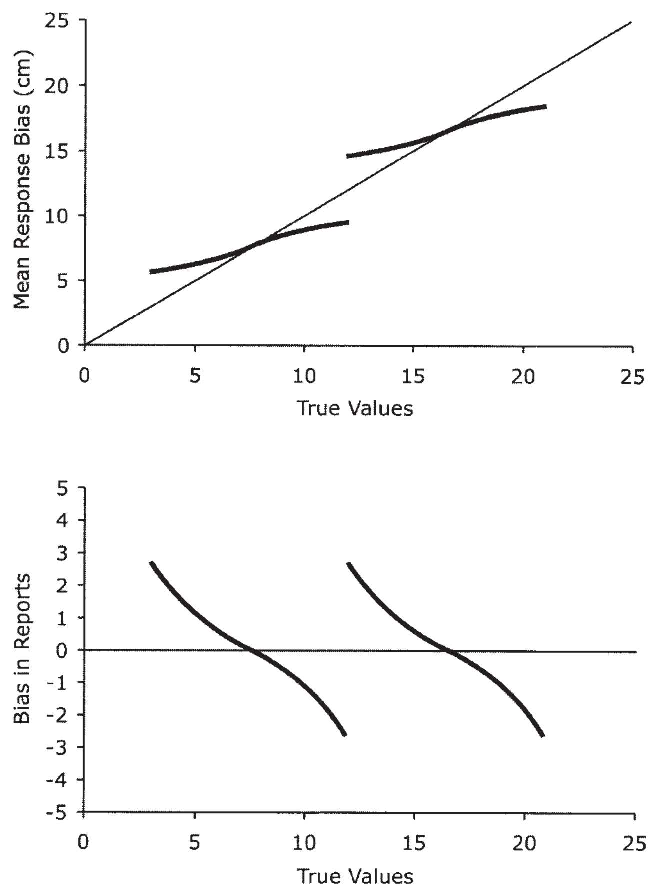

Figure 3. Graphs showing the effect of truncation at both lower and upper boundaries. The graph in the top panel shows the average reported values on the vertical axis plotted against the actual stimulus values on the horizontal axis. The pattern of bias from an upper boundary is the opposite of that from a lower boundary: it is negative, and bias is greatest for the stimulus values near the upper boundary (large stimuli). Again, if reported values were unbiased, they would lie on a 45° line. Instead, with truncation at both lower and upper boundaries, average reported values for smaller actual stimulus values lay above the 45° line, and average reported values for larger actual stimulus values lay below it. The graph in the bottom panel shows bias on the vertical axis. The bias is positive near the lower boundary and negative near the upper boundary.

probabilities of upper and lower truncation, is given by Equation 3 in Section 1 of the Appendix. This equation can be used to calculate the bias introduced by truncation as a function of stimulus inexactness, location relative to the boundaries, and probability of truncation at each boundary. Schematic bias curves based on hypothetical model parameters are shown in Figure 4, both for reported values (top panel) and for bias (bottom panel).

Assessment of the Truncation Model

The first step in our experimental work in this article was to establish whether observed bias shows the characteristic ("signature") pattern of truncation—large bias near boundaries that falls off as distance from the boundary increases. This signature bias pattern is highly distinctive, and its presence can be detected by visual examination of observed distributions of estimates. Although it is logically possible that there could be other models that would explain such a pattern, it is not obvious whether the observed bias pattern could be produced by a completely different mechanism or what that mechanism might be.

Having found the signature pattern of truncation in our first study (Experiments 1a & b), we tested a formal quantitative model in the second study (Experiments 2a & b). The model described in Section 1 of the Appendix is assessed via statistical inference. We used observed values to recover the unobserved distribution of fine-grain values and the probability of truncation due to boundaries, using the method of maximum likelihood estimation described in Section 2 of the Appendix. To gain some insight into how this estimation procedure works, consider an intuitive account of how the unobserved distribution for a particular stimulus near a category boundary can be recovered. We assumed the uncertain representation of the fine-grain value to be unbiased (i.e., centered at the true value) with values distributed symmetrically around that true value. Near a lower boundary, the upper portion of the distribution of stimulus uncertainty lies within the category and is not affected by truncation. Hence, it was possible for us to "fill in" (estimate) what the entire distribution of uncertainty around the true stimulus value would have been in the absence of truncation. Next consider how we recovered the likelihood of truncation for a stimulus. We counted the number of responses observed outside the boundary and compared it with the number that would have been expected for a symmetric underlying distribution for that stimulus in the absence of truncation.

The statistical procedures described in Section 2 of the Appendix exploit the intuitions just described. They yield estimates of the variance of the underlying distribution of fine-grain stimulus values (that would have been obtained if there had been no truncation) and of the likelihood that boundaries were used. We show in Section 3 of the Appendix that truncation is a Bayesian process and hence theoretically provides optimal estimates. Therefore, our model holds that truncation should improve accuracy across a wide variety of situations. To determine if the mental operations posited actually *do* improve accuracy, we can compare inferred estimates without adjustment with observed estimates with adjustment using the method described in Section 4 of the Appendix. That is, we could determine whether, despite the bias introduced by the truncation process, the estimates produced by the mediation process are still more accurate than the inferred unbiased fine-grained values.

There was an additional issue in assessing the truncation model in the context of multiple categories: Findings by Wedell (1995) and by Marks (1988, 1991) seem to contradict the category adjustment model. The results suggest that stimulus adjustment involves contrast rather than assimilation—that stimuli are adjusted toward boundaries and away from the category center. Recall that the two processes posited in our model—truncation due to boundaries and shrinkage toward a central value—both involve adjusting stimuli away from boundaries and toward the category center. We show that the apparent contrast findings would not arise if people used categories different from those that Wedell and Marks assumed that they used; then, the apparent contrast effects would be assimilation effects, as predicted by our model. (See section below.)

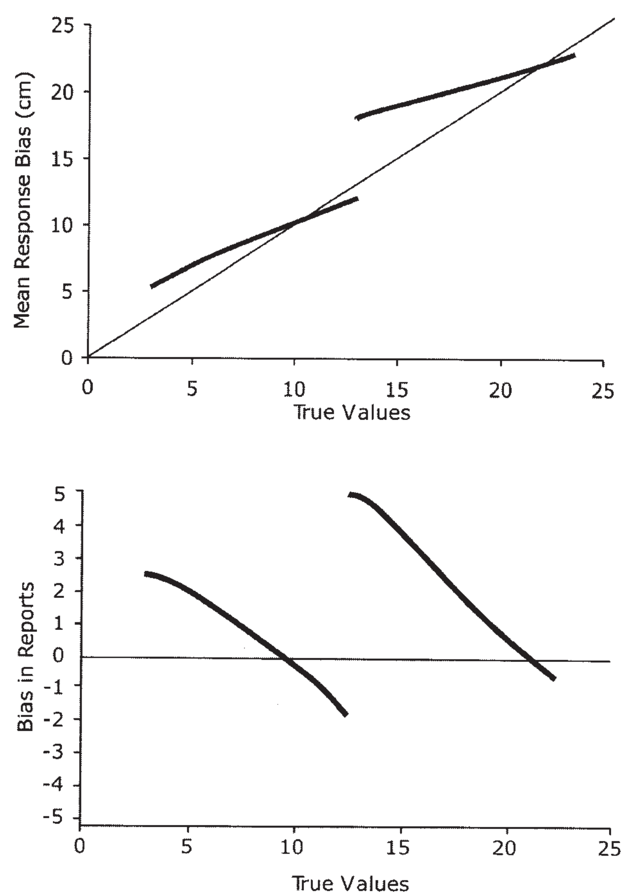

Figure 4. Effect of truncation at both lower and upper boundaries with varying boundary probabilities. The graph in the upper panel shows reported values, and the graph in the lower panel shows bias.

Earlier Work on Judgment of Stimuli From Contrasting Sets

Wedell (1995) and Marks (1988, 1991) obtained results that seem to show a contrast effect that would be inconsistent with our model. The condition in which the apparent contradiction arises is one in which there are two stimulus sets that vary in size and cover different ranges but partially overlap. In this situation, it has been found that equal stimuli in the region of overlap for the two sets are judged as unequal. The stimulus from the smaller set is judged to be larger than the stimulus from the larger set. Wedell (1995) presented two sets of squares distinguished by color (green or blue) and location (right or left). The stimuli in the two sets differed in size but overlapped partially (see Table 1). A pair of stimuli, one from each category, was presented simultaneously. The task involved a relative size judgment. For most pairs, the stimulus from the larger set was the larger one in the pair and was judged to be larger. Wedell also presented pairs that were of equal size for which it was not possible to judge one of the stimuli as bigger. People nevertheless judged the stimulus from the smaller category to be larger than the stimulus from the larger category. That is, the results indicate that estimates of the stimulus from the small blue category were moved away from the center of that category toward its upper boundary or that estimates of the stimulus from the large green category were moved away from the center of that category toward its lower boundary. Marks (1988, 1991) showed a similar contrast effect in a study involving judgments of the loudness of sounds.

On their face, Wedell's findings suggest that stimuli are adjusted away from a category center (contrast) rather than toward it (assimilation). In interpreting the findings, it should be noted that categories are groupings that a person has formed, and, as such, they should be distinguished from stimulus sets, which are groupings that an investigator has introduced. To present stimulus sets distinguished by cues such as color or location does not ensure that people will use those cues to form categories. In a classic study, Tajfel and Wilkes (1963) found that people subdivided stimuli that varied in length into two categories to correspond to a category cue. However, as McGarty (1999) has pointed out, several studies have failed to show subdivision of stimuli to correspond to such cues.

To determine which categories people are using, one must examine their responses. Wedell (1995) did not make such a test but rather assumed that people's categories were based on the color and location cues provided by the experimenter (e.g., blue on the left). If these were the categories people used, adjustment of stimuli would indeed be toward the boundaries of categories. However, there is another possibility. In Wedell's procedure, pairs of stimuli were presented together. When stimuli that vary only in size are presented simultaneously, people may code their relative size. Indeed, Wedell's task only required relative coding (larger or smaller), not absolute judgments (size estimates). If people made relative judgments, forming categories of "the larger square in a pair" and "the smaller square in a pair," thus ignoring color or location, there could be no overlap in the categories. Indeed, if categories were relative, observed bias in the region that overlapped in terms of color/location could be explained by our model. That is, adjustment toward the boundary of color/location categories would be toward the center of the relative categories—the larger square in a pair or the smaller square in a pair. This mechanism will become clear below.

Table 1
Stimulus Distributions From Wedell and Experiment 1b

Stimulus	Side lengths of squares in stimulus set			
	Small blue squares		Large green squares	
	Pixels	cm	Pixels	cm
1	70	2.50		
2	80	2.86		
3	90	3.21	90	3.21
4	100	3.57	100	3.57
5	110	3.93	110	3.93
6			120	4.29
7			130	4.64

THE EXPERIMENTS

To examine the estimation of stimuli (squares) that vary in size and that are divided into two sets distinguished by color/location cues, we presented two stimulus sets that differed in the range of sizes they encompassed, capturing a well-known situation in which conceptual or semantic domains are subdivided into distinct categories. We examined whether people form two separate categories by determining whether there are two distinct bias curves. If there were two curves, we determined whether the pattern of bias is that expected from truncation, exaggerating the differences between stimuli that cross category boundaries. In Experiment 1a, the squares were partitioned into two adjoining sets. In Experiment 1b, the two sets overlapped in size, allowing us to explore the source of the apparently paradoxical contrast effect found by Wedell (1995), as indicated below.

In the first experiment we used two conditions—single presentation and paired presentation. In Experiment 1a, we used adjacent stimulus sets. If color/location cues are the basis for categories, both forms of presentation would result in subdivision into two categories because the color/location cues of squares were available in both conditions. However, if relative size is the basis for categories, only paired presentation would result in subdivision into categories because the notion of relative size in a pair only arises for stimuli that are paired. Thus, if two categories are formed only for paired presentation, it would suggest that people use relative size categories.

In Experiment 1b, we used overlapping stimulus sets to examine further whether categories in the paired condition were based on color/location, as Wedell had assumed, or were instead based on relative size. The critical test concerns cases in the overlap region. The stimulus sizes in this region occurred in both categories. For color/location categories, some stimulus pairs in the overlapping region were anomalous in that the larger stimulus had the color/location of the smaller category or vice versa. For relative size categories, there was no such anomaly because whatever stimulus was larger in a presented pair was in the larger category. The evidence is that the categories used are relative size categories and that bias conforms to the category adjustment model. For both experiments (1a and 1b), the results indicate that truncation is the source of bias in estimation, but there is evidence of truncation only at the lower boundary.

The results of Experiment 1 led us to do a second study. Experiment 2 was designed to explore further whether truncation due to boundaries could explain patterns of bias in estimation and to model the results. We used a wider range of values for each stimulus set than in Experiment 1 to determine whether upper truncation as well as lower truncation was observed. We did this both with the two sets adjacent to one another (Experiment 2a) and with the sets overlapping (Experiment 2b). We tested a model with four parameters fitted from the data: the intercept and slope for fine-grain values, and the likelihood of using lower boundaries and/or upper boundaries to truncate extreme values.

For all four experiments in this article (1a and 1b, 2a and 2b), stimuli were presented and then were removed from the screen. Next, a response square in the same color and location as one of the squares was presented. Participants adjusted the size of this response square to reproduce the size of the square of that color and in that location that had just been presented. The response square used was very small, near zero, for two reasons. First, small stimuli are least inexactly represented. The response square was sufficiently small that the inexactness of the smallest test square did not overlap with it. Second, because zero is an absolute size limit, it was a natural truncation point. It is not artificially introduced by the experimental procedure. A large response square might have imposed an artificial upper boundary at that size point even if otherwise no upper boundary would have been used. Further, such a truncation effect due to an upper boundary would have been large because inexactness of stimulus representation increases with stimulus size. Finally, had we used a response square in the center of the size range, where color and location shift, we might have artificially produced a truncation effect at that point.

Experiment 1

Experiment 1a: Adjacent Stimulus Sets

In Experiment 1a, we examined estimation of stimuli from two adjacent sets that varied in magnitude. Each set was associated with a different color and location. There were two conditions—paired presentation, as in earlier work by Wedell (1995), and single presentation, as in our earlier work (e.g., Huttenlocher et al., 2000). The pattern of responses in our reproduction task could reveal what categories people form. Two distinct bias curves indicated the presence of two categories, and one bias curve indicated the presence of a single category. We anticipated that two categories would be formed for paired presentation but not for single presentation.

Method

Design. There were two presentation conditions. In the paired-presentation condition, a pair of squares was shown—one smaller blue square and one larger green square. In the single-presentation condition, only one square was shown. For each presentation condition, there were two possible stimulus positionings that were varied among participants; for a given participant, either the small blue squares appeared on the left and large green squares appeared on the right, or vice versa.

Participants. There was a total of 40 participants. There were 20 participants in the paired-presentation condition, with 10 in each group according to the positioning of stimuli (with target left vs. target right). There were 20 participants in the single-presentation condition, with 10 in each group according to the positioning of stimuli. All participants were current students at the University of Chicago.

Stimuli. We used stimuli based on the work of Wedell (1995). Stimuli were presented on a computer screen. For half of the participants in each presentation condition (single and paired), smaller blue squares were shown on the left and larger green squares on the right; for the other half of the participants, the smaller blue squares were shown on the right and the larger green squares on the left. Each stimulus set included squares of five different sizes evenly spaced. We described stimulus size as Wedell (1995) did—in terms of side length with pixels as units. For the smaller blue set, side length varied between 70 pixels and 110 pixels (2.5 cm and 3.8 cm). For the larger green set, side length varied between 120 pixels and 160 pixels (4.2 cm and 5.5 cm).

Procedure. There were 10 different stimuli to which responses were obtained for both the paired- and the single-presentation conditions. For the paired condition, there were 25 unique pairs, including all the pairs formed by pairing each member of one set with all the members of the other set. For each pair, there were 2 trials, involving each stimulus as target, for a total of 50 different types of trials. In the single-presentation condition, there were 10 unique trials including all squares, 5 trials from each stimulus set. For both paired and single presentation, each type of trial was conducted four times, for a total of 200 trials.

The participants' task was to adjust the response square to the estimated size of the stimulus square that was above it. In both conditions, the stimulus was presented for 1 s, and then it was removed from the screen. After 2 s, a very small response square (1 cm per side) appeared on the same side of the screen and in the same color as the square to be reproduced. The participants adjusted the size of the response square by pressing one of two vertically positioned computer keys—the top key to make the response square grow and the bottom key to make it shrink. Pressing the space bar registered the size of the response square.

Results

Responses from each participant for each stimulus were used to plot bias (the mean response size for a stimulus minus the true size for that stimulus). Note that in this experiment, the amount of observed bias indicated effect size. Size judgments were essentially identical regardless of presentation condition (small blue on the left and large green on the right, or vice versa) for both paired and single presentations. There were striking differences in response patterns for paired presentation versus single presentation. The top panel of Figure 5 shows two distinct bias curves for paired presentation, with intercepts that were significantly different, $t(7) = 16.56$, $p < .001$. The bottom panel shows only one bias curve for single presentation, with no significant difference in intercept between biases for the blue and green squares ($p > .05$). Thus, people formed two categories when stimuli were presented in pairs, one from each set, but formed only one category when stimuli were presented singly.

The bias curves can be examined visually for paired presentation (upper panel of Figure 5) and single presentation (lower panel of Figure 5). For paired presentation, bias within each category was most positive for values near the lower category boundary and fell off as distance from that boundary increased. The similarity of the graph in the upper panel of Figure 5 to the schematic graph in the lower panel of Figure 4 can be seen. For single presentation too

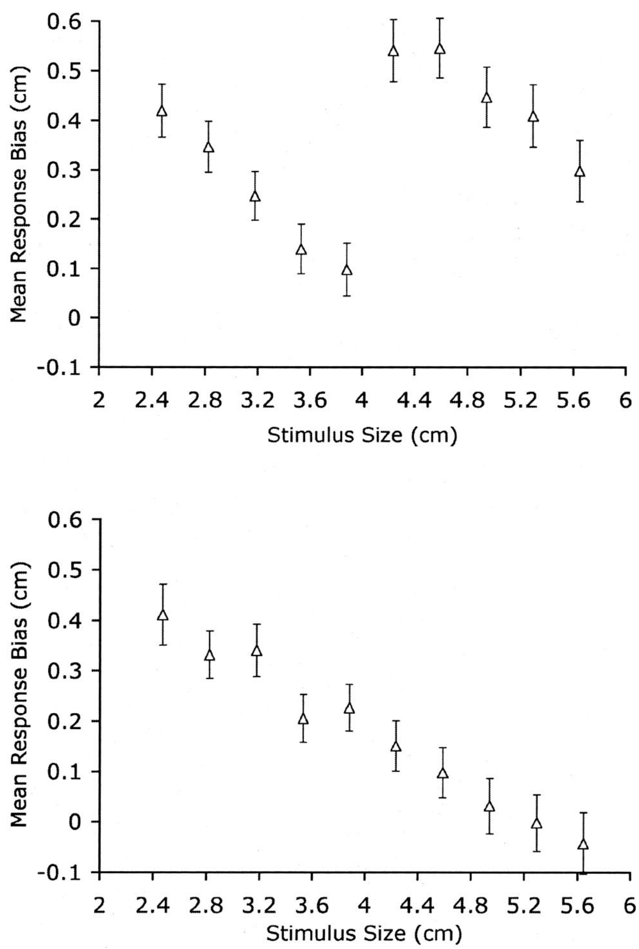

Figure 5. Response bias for adjacent categories in Experiment 1a. Mean response bias and stimulus size are plotted for paired presentation (upper panel) and single presentation (lower panel) of stimulus. There were two distinct bias curves for paired presentation: one curve for squares from the smaller stimulus set and one curve for squares from the larger stimulus set, with intercepts that were significantly different, $t(7) = 16.56$, $p < .001$. In contrast, the bottom panel shows that for single presentation, there was only one bias curve, with no significant difference in intercept between biases for the blue and green squares ($p > .05$). Triangle = mean value; bars = standard deviation.

(as shown in the lower panel of Figure 5), bias was greatest at the lower boundary and fell off as distance from that boundary increased. Recall that bias due to a lower boundary is always positive. So if lower boundary truncation is seen for all stimuli in a category, all the stimuli will be overestimated. The findings show that bias was always positive, with the greatest bias for values near the lower category boundary and less bias as distance from that boundary increased. Thus, the pattern of bias involved overestimation for all stimuli. This finding is consistent with truncation at the lower boundaries that extends across all values in the category.

Brief Discussion

We have found that two categories were formed for paired presentation but not for single presentation. This finding suggests that the categories are based on relative size rather than on color/location. If color/location were the basis of categorization (e.g., blue square on the left), one would expect categorization for both single and paired presentations. However, if color/location were ignored and categories were based on the relative size of present stimuli, categorization would only occur for paired presentation (e.g., large square in the presented pair). Yet in Experiment 1a, we could not directly distinguish relative size categories from color/location categories because for adjacent sets, there were no differential predictions about bias for these alternative forms of categorization. In Experiment 1b, we used partially overlapping categories in which, as indicated below, the expected results for the two forms of categorization can be differentiated.

Experiment 1b: Partially Overlapping Stimulus Sets

Experiment 1b was designed to be parallel to Wedell's (1995) study. As Wedell did in his study, we used two stimulus sets that overlapped partially, creating a region of overlap in which the sizes of the stimuli fell within the range of both stimulus sets. Wedell obtained apparently paradoxical findings with partially overlapping stimulus sets (i.e., contrast rather than assimilation). As indicated below, data from the region of overlap can provide information as to whether people use color/location categories or relative size categories.

The stimulus sets shown in Table 1 are the same in magnitude as the stimuli in Wedell's study. (Because most stimuli in one set were larger than those in the other, we refer to them as the larger and smaller stimulus sets.) Wedell used paired presentation and found that for equal size pairs, the stimulus from the smaller set was judged to be larger than the corresponding (equal size) stimulus from the larger set. This, he believed, reflected contrast—adjustment of one or both stimuli away from the center of the respective color/location category.

In terms of color/location categories, there were 25 distinct stimulus pairs, 22 of which were unequal. For 19 of these pairs ("consistent pairs"), the square from the larger category was larger in the presented pair than was the square from the smaller category. For 3 pairs ("anomalous pairs"), the relation was reversed, with the square from the larger set being smaller in the presented pair or with the square from the smaller set being larger in the presented pair. In terms of relative size, there were still 22 unequal pairs, but the 3 pairs that were anomalous for color/location categories were consistent for relative categories. That is, by definition, the larger stimulus in the pair was from the category "larger in the presented pair" and the smaller stimulus in the pair was from the category "smaller in the presented pair," regardless of color/location. Finally, there were 3 pairs in which the stimuli in the two sets were equal. Wedell found for equal size stimuli (at 90 pixels, 100 pixels, and 110 pixels), the stimulus from the small blue category was judged as being larger than the stimulus from the large green category, which was a contrast effect for color/location categories.

Method

Design and procedure. The design and procedure of Experiment 1b were the same as in Experiment 1a. Half the participants were in each of the presentation conditions (paired or single). For half of each group, the small blue square was on the left, and for the other, it was on the right. In two cases, the smaller blue set was shown on the left, and in two cases, the smaller blue set was shown on the right; for each, one had the stimulus from the smaller set as the target, and the other had the stimulus from the larger set as the target. Each of the pairs was shown twice, so the total number of trials was 200.

Participants. The experiment included 40 participants, 20 in each presentation condition (smaller blue on the left or right).

Stimuli. The stimuli in Experiment 1b were identical to those of Experiment 1a, except that the squares in the larger stimulus/distractor set were reduced in side length to overlap with the smaller stimulus/distractor set. The relation between the two stimulus sets, as described above, was identical to that of the sets used by Wedell (1995). Again, the stimulus size was measured in side length both in pixels (as in Wedell) and in centimeters (see Table 1). The smaller set varied from 2.5 cm to 3.8 cm, whereas the larger set varied from 3.2 cm to 4.5 cm. For single presentation, stimuli varied from 2.5 to 4.5 cm.

Results

Note that as in Experiment 1a, extent of observed bias was the effect size in this experiment. Visual examination of the bias curves for consistent pairs (the unfilled symbols in Figure 6) suggests that categorization was similar to that in Experiment 1a in that for paired presentation (upper panel of Figure 6), the bias was positive for all stimuli, with two distinct bias curves, one for each stimulus set. Bias fell off over all stimuli within each stimulus set. For single presentation also (lower panel of Figure 6), bias was positive and fell off over all stimuli; however, the responses form a single distinct bias curve. For paired presentation, we considered consistent and anomalous pairs separately to test whether participants used color/location or relative size in a pair in categorizing stimuli.

Paired presentation: Consistent pairs. The bias for pairs in which the square from the larger stimulus set was the larger square in the pair is shown as the unfilled symbols in the upper panel of Figure 6, revealing two categories. For these pairs (19 of 25 pairs), the figure shows two separate decreasing bias curves. Regression results confirm that these curves had significantly different intercepts, $t(7) = 7.03$, $p < .001$. The same pattern of results held for both presentation positions.

Paired presentation: Anomalous pairs. The category adjustment model predicts that smaller stimuli from the larger set should be adjusted upward and that larger stimuli from the smaller set should be adjusted downward. However, for color/location categories with overlapping stimulus sets, the size relations are reversed (anomalous) for some pairs—that is, the stimulus from the larger category is the smaller stimulus in the pair, or the stimulus from the smaller category is the larger stimulus in the pair (the filled symbols in Figure 6). If the category adjustment model with color/location categories were used with these anomalous pairs, filled and unfilled symbols of the same shape in Figure 6 should lie on top of one another. Instead we found a significant contrast effect for these anomalous pairs in which the predominant color/location pattern was reversed, $t(19) = 7.5$, $p < .001$.

Now consider the category adjustment model with overlapping stimulus sets in the context of relative size categories. For relative

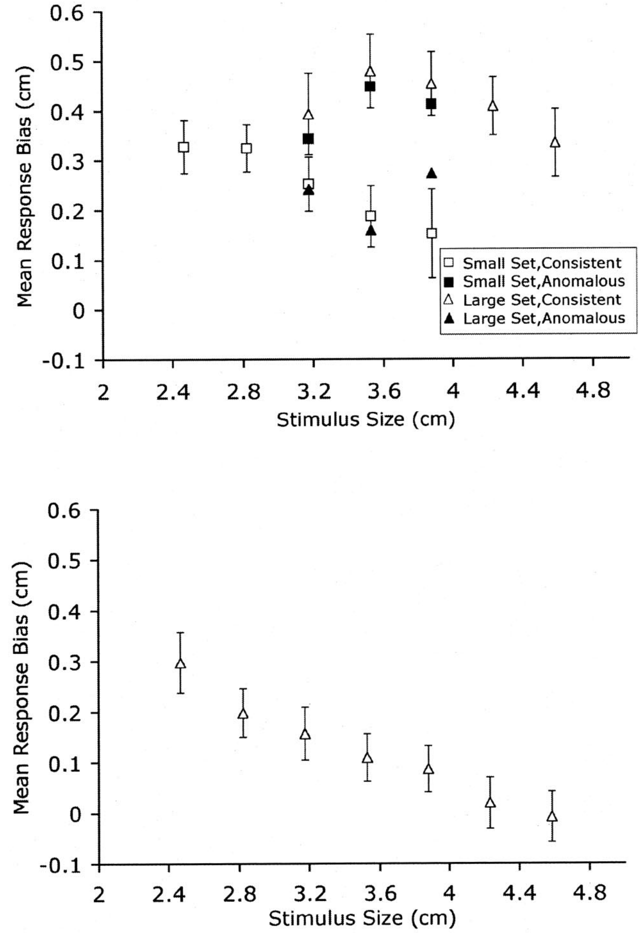

Figure 6. Response bias for overlapping categories in Experiment 1b. Upper panel shows paired presentation. Bias for pairs in which the square from the larger stimulus set was the larger square in the pair (consistent pairs) is shown as the unfilled symbols. Bias for pairs in which the stimulus from the larger category was the smaller stimulus in the pair or the stimulus from the smaller category was the larger stimulus in the pair (anomalous pairs) is shown as the filled symbols. Lower panel shows single presentation. Triangle/square = mean value; bars = standard deviations.

size categories, a smaller stimulus is a member of the smaller category, and a larger stimulus is a member of the larger category. Thus, stimuli from the larger set that would be anomalous for color/location categories are members of the category "smaller in presented set," and stimuli from the smaller set that would be anomalous in color/location categories are members of the category "larger in the presented set." If the category adjustment model is applied to relative size categories, bias for large-set consistent pairs should lie on top of small-set anomalous pairs and bias for small-set consistent pairs should lie on top of large-set anomalous pairs. What is a "contrast effect" for color/location categories is an "assimilation effect" for relative size categories. This is the pattern seen in Figure 6, indicating that people group stimulus according to whether the stimulus was the larger or the smaller stimulus in a presented pair.

Paired presentation: Equal pairs. Recall that Wedell found that for equal size stimuli, the square from the small set was judged to be larger than that from the large set and that the square from the large set was judged to be smaller than that from the small set (contrast). Our results for equal size stimuli were in the same direction as Wedell's results. However, for equal size stimuli from the different sets, the differences were not statistically significant (all $ps > 0.05$). Nor was the difference significant in either presentation order. When stimuli consisted of pairs of the same two sizes, the bias observed was the same, regardless of color/location.

Single presentation. For single presentation, plotting average bias for each stimulus against actual stimulus size, collapsed across presentation orders, revealed only one bias curve, shown in the lower panel of Figure 6. The bias curve was very similar to that for single presentation in Experiment 1a. The difference in intercept between biases for the blue and green squares did not reach significance ($p > .05$). The same pattern of results held for both presentation orders. As in Experiment 1a with adjacent sets, Experiment 1b with overlapping sets showed that for single presentation, people formed one category, and for paired presentation, they formed two categories. For both single and paired presentations, there was overestimation of stimuli. The overall bias pattern for both single and paired presentations was consistent with truncation at the lower boundary.

Brief Discussion

In Experiments 1a and 1b, we found that the bias in stimulus estimation for a pair of contrasting categories was a truncation effect. However, evidence of truncation was found only at lower boundaries. Although extreme low values that fall outside category boundaries were truncated, we found no evidence of such a process at upper boundaries. This might be because the categories were so narrow that lower truncation still affected estimates even for the largest values in the category. In addition, we obtained further evidence that people used relative size categories. They only subdivided stimuli into two categories when the stimuli were presented in pairs. In addition, in Experiment 1b, we showed how Wedell's finding of a contrast effect with overlapping sets could be interpreted as an assimilation effect if people used relative size categories.

Experiment 2: A Truncation Model of Estimation for Contrasting Stimulus Sets

In Experiment 2, we modeled truncation effects with paired presentation. Because we were concerned with truncation for contrasting categories and because Experiment 1 showed that people form two categories only with paired presentation, we used only paired presentation in Experiment 2. The design was the same as that in the paired conditions in Experiments 1a and 1b, except that stimuli within each set covered a wider range of values, 10 versus 5 values. Our aim was to determine whether there was evidence of upper as well as lower truncation with wider stimulus sets. In Experiment 2a, the two stimulus sets were adjacent, and in Experiment 2b, the two stimulus sets overlapped. We explored whether truncation actually increased average accuracy as posited in the category adjustment model.

We tested a mathematical model in which truncation of extreme values accounted for the pattern of bias in stimulus estimation. We fitted the data from each experiment (Experiments 2a and 2b) to the model. We based the assumptions about stimulus representation in the truncation model on the category adjustment model, which holds that encoding of both fine-grain and category information is inexact but unbiased. We assumed that the inexactness of stimulus coding would increase linearly with stimulus magnitude, as in Weber's law. The function was captured in two parameters in our model, an intercept and a slope, which were fitted from the data of an experiment.

Boundary locations were set at their true values—the extreme values of each of the stimulus sets. However, because boundary location is derived from encounters with stimuli, it is uncertain. We set this uncertainty at 50% of the mean uncertainty of particular stimuli. Our reasoning was that information about boundary locations is based on accumulating exposures to the extreme stimuli in a set (smallest and largest), so the inexactness of boundary location would be less than that for a single stimulus. The likelihood that truncation may be used could differ for lower and upper boundaries. We used the experimental data to estimate the likelihood of participants' using lower and upper boundaries in that experiment. As we have noted, upper boundaries may be used less often than lower boundaries. However, we assumed that the likelihood of participants' using lower boundaries is the same in smaller and larger categories and that the likelihood of their using upper boundaries also is the same in smaller and larger categories. In short, the model that we tested had four parameters that were fitted from the data. Two had to do with stimulus uncertainty (intercept and slope), and the other two had to do with the likelihood of using truncation at lower and upper boundaries. These parameters were firmly based on the logic and findings of the category adjustment model. The values for the parameters can be expected to differ with particular experimental conditions.

Experiment 2a: Model for Adjacent Stimulus Sets

In Experiment 2a, we tested whether the model could account fully for the patterns of bias observed for adjacent stimulus sets. The parameters were fitted from the data to test the model. To further evaluate the goodness of fit of the model, we estimated parameters from one half of the participants and tested whether these parameters fitted the data from the other half of the participants. We examined whether the use of truncation increased average accuracy of estimates, as the model predicts.

Method

Design. As in the paired condition of Experiment 1a, a pair of squares was shown—the smaller blue square on one side and the larger on the other side (e.g., small on the left and large on the right), with the stimulus positioning counterbalanced across participants. For a given participant, either the smaller blue squares appeared on the left and larger green squares on the right, or vice versa.

Participants. There were 40 participants, 20 in each group with a particular positioning of smaller blue versus larger green squares. All participants were drawn from the University of Chicago community.

Stimuli. The two stimulus sets for the estimation task included 10 smaller blue squares on the left (or right) and 10 larger green squares on the right (or left), displayed on a computer screen. In

Experiment 2, in contrast to Experiment 1, stimulus size was presented in square centimeters. In Experiment 1, we used only side length because we were testing the ideas presented by Wedell (1995) who used side length as the measure. However, for the mathematical modeling carried out in Experiment 2, we used square centimeters because area may better capture the coding involved. The stimulus sets were evenly spaced from an area of between 5 cm^2 and 50 cm^2 in the smaller blue set and from an area of between 55 cm^2 and 100 cm^2 in the larger green set in increments of 5 cm^2.

Procedure. On each trial, a square from the smaller set of 10 was paired with a square from the larger set of 10. Thus, there were 100 unique pairs, each shown twice, because the participants estimated the size of both the smaller and larger squares in the pair. The total number of trials was 200. The trials were completely randomized. In Experiment 2, the response was made via a tracking ball rather than computer keys. The participants could rapidly decrease the size of the response square by moving the tracking ball to the left and increase it rapidly by moving it to the right. Pressing the space bar registered the area of the response square.

Results

Responses from each participant for each stimulus were used to plot bias (Figure 7), which is the effect size in this experiment. In plotting the pattern of bias, we used relative categories because we found in Experiment 1 that these are the categories that people use. We found a distinct curve corresponding to each category—smaller square in a pair and larger square in a pair. We used a simple linear model to test for the presence of a category effect and found that the intercepts of the bias curves differed significantly, $t(17) = 15$, $p < .001$. The same pattern of results held for both presentation positions, with no evidence that position affected estimation. The biases of estimates were entirely positive and decreased in a pattern similar to those for the categories in the paired condition of Experiment 1.

Table 2 shows the parameters that resulted from fitting the model to participants' responses: for the slope and intercept of the line relating stimulus size to memory uncertainty and for the probabilities of lower and upper truncation. We assessed the fit of the model using 95% confidence intervals for the difference between observed and predicted means. As shown in Figure 7, observed means fell within the 95% confidence intervals for all 20 stimuli, 10 in each category. The correlation of predicted and observed bias was .97. There was evidence of use of an upper as well as a lower boundary for this experiment, in which the set of stimuli was wider than the corresponding set in Experiment 1. However, participants were less likely to use the upper than the lower boundary. The probability of truncation at a lower boundary was close to 1.00 (0.982), whereas the probability of truncation at an upper boundary was 0.497.

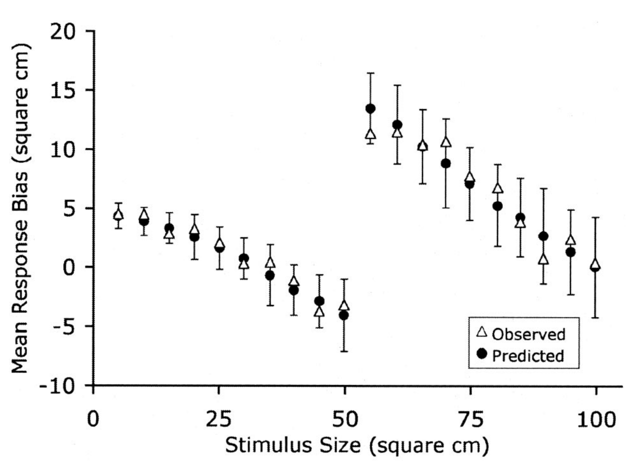

Figure 7. Predicted and observed response bias for adjacent categories in Experiment 2a with a combination of lower and upper boundaries. Bars = standard deviations.

Table 2
Fitted Model Parameters for Bias in Area Estimates: Adjacent Stimulus Sets, Experiment 2a

Parameter	Estimate	*SE*
Slope of memory uncertainty	0.297	4.81×10^{-4}
Intercept of memory uncertainty	6.381	2.15×10^{-2}
Probability of lower truncation	0.982	4.58×10^{-6}
Probability of upper truncation	0.497	5.37×10^{-6}

To determine whether the excellent fit of the statistical model to the data could, in part, be due to chance factors, we used a cross-validation method—that is, by estimating the model with only one part of the data and then assessing the model using the other part of the data. We randomly drew half of the observations at each of the 20 stimulus values and fitted the model to these observations. We then assessed the fitted model on the remaining data—the other half of the observations—by comparing the observed bias to the bias predicted by the model. The fit was as good as that for the entire dataset. Observed mean values fell within the 95% confidence intervals for the predictions for 19 of 20 stimuli, 9 in the smaller stimulus set and 10 in the larger stimulus set. Thus, the model makes accurate predictions that are not due to overfitting of a particular configuration of data.

As can be seen in Section 4 of the Appendix, we tested whether truncation would improve the accuracy of estimation, even if truncation were used only some of the time. We considered the impact of truncation when boundaries were used with various probabilities by conducting simulations based on the memory uncertainties inferred from the data. We did this for a variety of different probabilities of truncation at lower boundaries (indicated on each row) and at upper boundaries (indicated on each column). Table 3 gives the proportional reduction in mean square error, compared with a model in which no truncation was used. The value corresponding to a probability of lower boundary truncation of 0.98 and an upper boundary truncation probability of 0.50 is given in boldface in the table because they were the values estimated from the data in Experiment 2a. They indicate that the use of the values estimated in Experiment 2a reduced the error of estimation by approximately 64%. More generally, for a variety of fitted models, truncation with probabilities of use that were less than 100% still resulted in accuracy greater than that for unadjusted estimates. Clearly, the use of truncation improves accuracy in a wide range of cases.

Table 3
Ratio of Mean Square Error With Truncation (at Lower Boundary With Probability p_a *and at Upper Boundary With Probability* p_b*) to Mean Square Error With No Truncation*

	p_b													
p_a	0.00	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.98
0.00	1.000	0.999	0.997	0.995	0.990	0.985	0.980	0.976	0.972	0.970	0.969	0.971	0.973	0.976
0.01	0.998	0.997	0.995	0.993	0.988	0.983	0.978	0.973	0.970	0.967	0.966	0.967	0.970	0.972
0.05	0.989	0.989	0.986	0.984	0.978	0.973	0.967	0.962	0.958	0.955	0.953	0.953	0.955	0.957
0.10	0.978	0.978	0.975	0.972	0.966	0.960	0.954	0.949	0.944	0.939	0.936	0.935	0.937	0.938
0.20	0.956	0.955	0.952	0.949	0.942	0.935	0.928	0.921	0.914	0.907	0.902	0.899	0.898	0.899
0.30	0.933	0.932	0.929	0.925	0.917	0.909	0.900	0.892	0.883	0.875	0.867	0.861	0.859	0.858
0.40	0.909	0.909	0.905	0.900	0.891	0.882	0.872	0.862	0.851	0.841	0.831	0.821	0.817	0.815
0.50	0.885	0.884	0.880	0.875	0.865	0.854	0.843	0.831	0.818	0.806	0.793	0.780	0.774	0.770
0.60	0.860	0.859	0.855	0.849	0.838	0.825	0.813	0.799	0.784	0.769	0.754	0.737	0.729	0.723
0.70	0.835	0.834	0.829	0.823	0.810	0.796	0.781	0.766	0.749	0.732	0.713	0.693	0.682	0.674
0.80	0.809	0.807	0.802	0.795	0.781	0.766	0.749	0.732	0.713	0.693	0.670	0.646	0.633	0.623
0.90	0.781	0.780	0.774	0.767	0.751	0.734	0.716	0.697	0.675	0.652	0.626	0.597	0.582	0.569
0.95	0.768	0.766	0.760	0.752	0.736	0.718	0.699	0.679	0.656	0.631	0.603	0.572	0.555	0.540
0.98	0.756	0.755	0.749	0.741	0.724	0.705	0.686	**0.664**	0.640	0.614	0.585	0.552	0.533	0.517

Note. Bold value = the setting for upper truncation probability and boundary uncertainty that was most similar to the fitted model in Experiment 2.

Brief Discussion

Our truncation model with four parameters accounted for the bias in responses to stimuli from the smaller and larger stimulus sets. It should be noted that two of these parameters are not free in the typical sense; rather, they describe well-known cognitive processes involved with the inexactness of representation that vary from one context to another. The other two parameters are descriptive, indicating the likelihood that boundaries would be used in estimation. The results indicated that participants formed two categories with boundaries that truncated extreme sampled stimulus values. The fitted model indicated that upper as well as lower truncation, used with different probabilities, was important in accounting for the bias in responses to larger values, even in the larger stimulus set in which there was no underestimation. The adjustment improved average accuracy.

Experiment 2b: Model for Partially Overlapping Stimulus Sets

In Experiment 2b, we extended our evaluation of the formal model that we have proposed. In particular, we tested the truncation model with overlapping stimulus sets similar to those in Experiment 1b but with 10 stimuli in each set, as in Experiment 2a. The partially overlapping sets are shown in Table 4. We assumed that people use relative categories as we showed they did in Experiment 1b. Consider the distribution of stimuli for the categories of "smaller stimuli in a pair" and " larger stimuli in a pair." The stimuli varied from 5 cm to 75 cm. The range of sizes for the set "smaller stimuli in a pair" included 10 target stimuli from 5 cm to 50 cm. The range of sizes for the set "larger stimuli in a pair" included 10 target stimuli from 30 cm to 75 cm. Stimuli from 5 cm to 25 cm were always in the smaller set; stimuli from 55 cm to 75 cm were always in the larger set. Values from 30 cm to 50 cm fell in the overlap region so that target stimuli could have been in the smaller or the larger category. There were five cases in which stimuli in the pair were equal in size.

Method

Design. Squares were presented in pairs, one square from the smaller set and one from the larger set. The positions of the smaller blue and larger green squares were varied; for half the participants, smaller blue square was on the left, and for the other half, smaller blue square was on the right.

Participants. There were 40 participants with 20 in each group according to the positioning of the stimuli. All participants were students at the University of Chicago.

Stimuli. The positions of the two stimulus sets were varied across participants. For half the group, smaller blue squares were on the right and larger green sets were on the left, and for the other half the positions were reversed. The stimuli were pairs of squares. The squares in the smaller blue set were of 10 different sizes, evenly spaced between 5 cm^2 and 50 cm^2 in increments of 5 cm^2.

Table 4
Stimulus Values for Experiment 2b

	Area of squares in stimulus set (in cm^2)	
Stimulus	Small blue squares	Large green squares
1	5	
2	10	
3	15	
4	20	
5	25	
6	30	30
7	35	35
8	40	40
9	45	45
10	50	50
11		55
12		60
13		65
14		70
15		75

The squares in the larger (but overlapping) green set were similarly of 10 sizes, evenly spaced between 30 cm^2 and 75 cm^2, also in increments of 5 cm^2.

Procedure. On each trial, a stimulus from the blue (smaller) set was paired with a stimulus from the green (larger) set. Again, the total number of trials was 200. As in Experiment 2a, participants used a tracking ball to respond.

Results

In Figure 8, we used the responses from each participant to construct the bias curves shown; these are the effect sizes for the experiment. Examination of the bias curves shows the presence of two categories: smaller in a pair, larger in a pair. We used a simple linear model to test for the presence of a category effect and found that the intercepts of the bias curves differed significantly, $t(17) = 26.3$, $p < .001$, indicating that the participants used two categories. The bias of estimates was entirely positive and decreased as in the paired condition in Experiment 1b and in Experiment 2a. There were five equal size pairs from the overlap region (from 30 cm^2 to 50 cm^2). Recall that squares of these sizes in this region belonged to both small and large stimulus sets. If a target was blue (from the smaller set), then according to Wedell's findings, it would be moved toward the upper boundary of that set. If it was green (from the larger set), it would be moved toward the lower boundary of the larger set. As for equal pairs in Experiment 1b, differences between estimates were not statistically significant as a function of whether the target was from the blue or green set (all ps $> .05$).

Table 5 shows the parameters that resulted from fitting the model to participants' responses; the slope and intercept relating stimulus size to memory uncertainty and the probabilities of lower and upper truncation. We assessed the fit of the model using 95% confidence intervals for the difference between observed and predicted means. As shown in Figure 8, observed means fell within the 95% confidence intervals for all 20 stimuli. The correlation of predicted and observed means was .98. The probability of truncation fitted for the lower boundary was 0.971 and for the upper boundary was 0.299. Thus, even though bias was always positive, there was evidence of use of an upper as well as a lower boundary. We did calculations similar to those made for Experiment 2a (and shown in Table 3) in which we showed that the use of truncation at lower and upper boundaries across a broad range of probabilities improved overall accuracy compared with estimates in which truncation was not used.

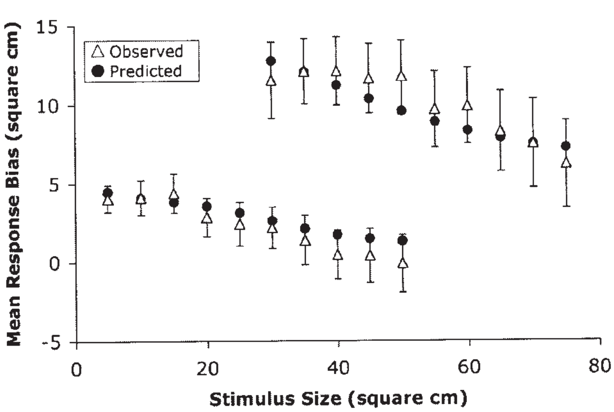

Figure 8. Predicted and observed response bias partially overlapping categories in Experiment 2b with a combination of lower and upper boundaries. Bars = standard deviations.

Table 5
Fitted Model Parameters for Bias in Area Estimates: Overlapping Stimulus Sets, Experiment 2b

Parameter	Estimate	*SE*
Slope of memory uncertainty	0.562	3.73×10^{-2}
Intercept of memory uncertainty	2.325	4.5×10^{-3}
Probability of lower truncation	0.971	2.22×10^{-2}
Probability of upper truncation	0.299	1.47×10^{-2}

Brief Discussion

As in Experiment 2a, our truncation model with upper and lower boundaries fully accounted for the bias in responses to stimuli from the larger and smaller stimulus set. Again, let us note that the parameters in our model are not free in the ordinary sense but rather describe well-known cognitive processes. Fitting the model to the data indicated that both lower and upper boundaries were used, but that lower boundaries were used with a higher probability than were upper boundaries. Thus, although the bias was positive throughout, the model showed that the data fit best if use of an upper as well as a lower boundary is assumed. As in Experiment 2a, simulations showed that truncation increased average accuracy across a broad range of probabilities.

GENERAL DISCUSSION

This article concerns people's use of inexact information in memory to estimate characteristics of events they have encountered. Information about events in memory is inexact, yet knowledge of the past is important in responding properly in the present. We described procedures by which people increased the accuracy of estimates of the what, when, and where of events under conditions of uncertainty. The general principle we explored involves combining inexact information from multiple sources to increase accuracy. We have examined one application of this principle: treating an inexactly represented stimulus in the context of a category that supplements memory for a particular target stimulus by summarizing prior experiences that specify a range and distribution of instances. The category indicates which stimulus values are most likely to occur. Combining category information with fine-grain information about a particular stimulus results in bias for individual estimates. Nevertheless, it can increase average accuracy by decreasing the availability of estimates. Although category adjustment processes are ubiquitous, people generally are unaware of using them; they believe that they simply report what they remember.

Category Adjustment: The Use of Truncation

The category adjustment process we have explored in the present article is truncation, a procedure that constrains stimulus estimates to fall within category boundaries. In the case we explored, people estimated stimuli from contrasting categories. The boundary of separation between categories provides important information in such cases (e.g., whether mushrooms are poisonous or are edible or whether insects bite or do not bite). Goldstone (1996) found that in such situations, there may be substantial exaggeration of the differences between stimuli from different categories. In our experiments, we found large bias effects in a task in which people estimated the sizes of squares in different stimulus sets. We found substantial exaggeration of intercategory differences due to truncation in estimation. Although bias at category boundaries was large, it could be fully explained by our model, which posits truncation at boundaries. Using a formal mathematical model, we showed that the use of truncation improved average accuracy over unadjusted estimates in a wide range of cases.

The truncation effect that we found was most striking at lower boundaries. One factor in this asymmetry concerns the nature of magnitude scales. Magnitude is bounded by zero at the lower end and is unbounded at the upper end. The dimension of size is one of bigness, not smallness, so a category might be defined as "bigger than *X*," where *X* is a lower boundary, not "smaller than *Y*." Further, larger items are coded less exactly than smaller items, so truncation at the lower boundary of the larger set produces greater upward bias than for the smaller set, exaggerating the difference between stimuli at the boundary separating the two categories.

Further insight into the structure of categories that encode quantity is found in the lexical meanings of quantitative adjectives. There is a set of words that show the salience of coding of relative size for present objects. When a target object is presented alone, relative adjectives, such as tall, long, and big, are generally used for size relative to the mean of the category (e.g., "big dog" means big relative to the category of big dogs). However, when two objects are presented simultaneously, the word "big" (in the term "big dog") generally refers to the size of a target object relative to the other present object, even when both objects are small relative to the category mean. An intuition about the importance of relative size coding is that it is important in the use of language to individuate stimuli—for example, "give me the big one." The latter sort of relativity is involved in categorization in the present study.

Improving Estimation by Using Prior Information

Not only are particular memories inexact, but category information also is inexact. Although knowledge of inductive categories is never complete, simplifying assumptions can be used to approximate prior distributions—boundaries, central value, and dispersion of instances. Our previous research concerned the use of a central category value (prototype) and dispersion (e.g., Huttenlocher et al., 1991, 2000). If properly weighted with an inexact fine-grain value, the mean and dispersion of instances provide optimal information about particular stimuli. We have shown that relative to ignoring category information, one can improve accuracy by adjusting toward any fixed value (Huttenlocher, Hedges & Vevea, 2000).

In the present article, we have shown that truncation of category boundaries can improve accuracy. The procedure of truncating values at boundaries can increase accuracy for a wide range of adjustments that are not optimal but nonetheless improve accuracy compared with use of unadjusted fine-grain values, as shown in the Appendix. That is, when boundaries are uncertain or when truncation is not used 100% of the time, truncated estimates are still more accurate than uncorrected estimates. Although improvement in accuracy over uncorrected estimates is greater when boundaries are more precise and the probability of truncation is higher, truncation is useful except when the uncertainty of the boundaries is very high (i.e., the point at which uncertainty approaches the uncertainty of particular memories) or when the probability of truncation is very low.

Relative Size Categories

An important feature of the category adjustment model is that it makes it possible to infer the nature of people's categories. Using the model, one can reason backwards from observed biases to the categories that may have led to those biases. This is important because categories are not directly observable; they must be inferred from people's behavior. When there are many ways to group a particular set of stimuli, convincing evidence about which categories people use may be difficult to obtain.

In the present study, we provided evidence of the use of categories that are not frequently examined in the category literature, namely those that code relative quantity. In our experiments, we presented squares that varied over a range of sizes. The stimuli were divided into two subsets characterized by different category cues (color/location cues). The size ranges for these subsets were either adjacent, or they overlapped partially. It seemed a likely assumption that people's categories would specify the range of values associated with a particular category cue (i.e., stimuli falling in the range from 5 cm^2 to 50 cm^2 [blue on left] and stimuli falling in the range from 55 cm^2 to 100 cm^2 [green on right]). However, when Wedell presented stimuli in pairs, one from each subset, with categories that overlapped partially, the observed biases could not be accommodated by the category adjustment model. The problematic finding was that values in the overlap region were adjusted away from category centers and toward boundaries. This possible anomaly can be resolved if one posits that people use relative size categories defined as the larger (or smaller) stimulus in a presented pair. In that case, people's responses to stimuli in the overlap region would involve bias toward the center and away from the boundaries. We were able to model truncation for relative size categories with both adjacent and overlapping pairs. Providing further evidence that people found categories that were relational was the finding that when squares were presented singly, people did not subdivide them into two subsets at all. They used relative size categories only when stimuli from two stimulus sets were paired in presentation.

To find that the categories people used were based on relative size is important both because it shows how the nature of the categories people use can be inferred from the patterns of bias in responding and also because relative categories are an important kind of stimulus grouping. As Goldstone (1996) has noted, it is possible to distinguish between two sorts of categories. In one sort, categories are independent of one another, each capturing a particular range of stimulus values (e.g. stimuli of a size between 30 cm^2 and 50 cm^2). In the other sort, categories are interdependent, capturing relative information (e.g., stimuli that are smaller in presented pairs). Such interdependence characterizes some lexical categories but has not been systematically examined for nonlanguage categories like those in the present experiments.

Other Cases of Information Combination

Although using categories to adjust inexact fine-grain values is an important way to combine multiple information sources, there are cases in which using multiple sources can improve accuracy, for example, in determining the location of an object (relating it to both distal and proximal landmarks) or in determining the time of an occurrence (relating it to both the date and time that has elapsed from the present). That is, combining information from different sources does not always involve hierarchically organized information. If the information from a single source were exact, combining sources could not improve accuracy. However, as we have noted, information in memory is rarely exact. An important principle in combining information from different sources is to weight the information according to its precision, giving more weight to information that is more precise.

There is evidence that humans combine visual and haptic information in a nearly optimal fashion (e.g. Ernst & Banks, 2002; Gepshtein & Banks, 2003). Ernst and Banks asked study participants to judge the width of stimuli that were presented visually, haptically, or bimodally. The variance of responses based on unimodal stimulus information was greater than the variance of responses based on combined bimodal information. Vision is often more reliable than haptic information, and in that case, greater weight should be given to that information than to haptic information, which is more variable. However, when visual information is made less reliable, people place greater weight on the haptic information. Indeed, the weights given to the two information sources were near optimal, maximizing the accuracy of estimates of object height. Combining different sources of information from the same modality has been shown to improve accuracy; for example, stereo and texture cues in the visual modality can be combined to increase accuracy (Knill & Saunders, 2003). In short, the principle of improving accuracy by combining information can involve a wide range of cases in which there are different pieces of inexact information that pertain to the same stimulus.

A stimulus estimate then is a blend in which greater weight is given to the more exact source. When information about particular stimuli is made more variable via an interference task, people will weight category information more heavily (see Huttenlocher et al., 1991). Further, when category information is more variable because instances are more dispersed (uniform rather than normal), people weight the stimulus more heavily (see Huttenlocher et al., 2000). This principle has been extended here in explaining the exaggeration of differences between two adjacent or partially overlapping categories.

Conclusions

In conclusion, we have proposed and tested a category adjustment model that posits transformation processes that mediate between the underlying representation of stimuli and of categories (both of which are unobservable) and stimulus estimates (which are observable). The transformation processes integrate prior information with a fine-grain value according to Bayesian principles. Because stimulus representations are not directly observable, the model uses observed stimulus estimates to infer them. Starting with the assumption that fine-grain stimulus representation is unbiased, with an unknown variance, use of the observed stimulus estimates allows us to deduce this variance from the observed stimulus estimates using maximum likelihood estimation (see, e.g., Rao, 1965). By computing the accuracy of the unobserved fine-grain representations using observed estimates, one can evaluate whether the transformation processes we posited increase accuracy compared with the untransformed fine-grained values. We examined the extent to which the processes embodied in a truncation model increased accuracy for a range of parameter values including (but not limited to) those estimated from the experimental data. The process increased accuracy across a wide range of values.

References

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

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

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

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

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.

Ernst, M. O., & Banks, M. S. (2002, January 24). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433.

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.

Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology, 13,* 483–488.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory and Cognition, 24,* 608–628.

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., & 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.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129,* 220–241.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43,* 2539–2558.

Johnson, N. L., & Kotz, S. (1970). Distributions in statistics: Continuous univariate distributions—1. New York: Wiley.

Marks, L. E. (1988). Magnitude estimation and sensory matching. *Perception and Psychophysics, 43,* 511–525.

Marks, L. E. (1991, April). *Slippery context effects in psychophysics*. Paper presented at the meeting of Experimental Psychology Society, Cambridge, England.

McGarty, C. (1999). *Categorization in social psychology.* London, England: Sage.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Poulton, E. C. (1989). *Bias in quantifying judgments.* Hillsdale, NJ: Erlbaum.

Quinton, A. (1957). Properties and classes. *Proceedings of the Aristotelian Society, 48*(Suppl. vol.), 33–58.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika, 52,* 447–458.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Sudman, S., & Bradburn, N. M. (1973). Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association, 63,* 805–815.

Tajfel, H. (1959). Quantitative judgment in social perception. *British Journal of Psychology, 50,* 16–29.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. *British Journal of Psychology, 54,* 101–114.

Wedell, D. H. (1995). Contrast effects in paired comparisons: Evidence for both stimulus-based and response-based processes. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1158–1173.

Appendix

Experiment 2 Truncation Model: Procedure and Proof

In this appendix, we describe the truncation model that is used with minor variations in Experiments 2a and 2b. We also describe the maximum likelihood estimation (MLE) procedure used to estimate parameters in the truncation model. Then we prove the theoretical result that classical truncation (truncation at a boundary with a probability of 1) is a Bayesian procedure, and thus it increases accuracy. Finally, we show empirically that in our truncation model, in which truncation occurs with probabilities of less than 1, truncation still increases accuracy of estimation.

Section 1: The Truncation Model in Experiment 2

A truncated estimate Y for a stimulus μ_O, based on an unbiased underlying memory with standard deviation σ and certain lower and upper boundaries a and b, has the following distribution:

$$f(y|\mu_0,\sigma) = \frac{\left[\phi\left(\frac{y-\mu_0}{\sigma}\right)/\sigma\right] I_{a,b}(y)}{\Phi\left(\frac{b-\mu_0}{\sigma}\right) - \Phi\left(\frac{a-\mu_0}{\sigma}\right)} \quad (1)$$

(see, e.g., Johnson and Kotz, 1971). When the boundaries are assumed to be inexact about means μ_a and μ_b with standard deviations σ_a and σ_b, the distribution of the response Y no longer has a simple form, but the probability density function can be expressed as follows:

$$f(y|\mu_0,\sigma,\sigma_a,\sigma_b) = \frac{\left[\phi\left(\frac{y-\mu_0}{\sigma}\right)/\sigma\right]\Phi\left(\frac{y-\mu_a}{\sigma_a}\right)\left[1-\Phi\left(\frac{y-\mu_b}{\sigma_b}\right)\right]}{\int_{-\infty}^{\infty}\left[\phi\left(\frac{s-\mu_0}{\sigma}\right)/\sigma\right]\Phi\left(\frac{s-\mu_a}{\sigma_a}\right)\left[1-\Phi\left(\frac{s-\mu_b}{\sigma_b}\right)\right]ds}. \quad (2)$$

Finally, if left or right truncation occurs only with probabilities p_a and p_b, the distribution of the response Y is a mixture of such truncated distributions with a normal distribution centered at the true stimulus mean with standard deviation given by the memory uncertainty.

$$\begin{aligned} f(y|\mu_0,\sigma,p_a,\sigma_a,p_b,\sigma_b) &= (1-p_a)(1-p_b)\left[\phi\left(\frac{y-\mu_0}{\sigma}\right)/\sigma\right] \\ &+ p_a(1-p_b)\frac{\left[\phi\left(\frac{y-\mu_0}{\sigma}\right)/\sigma\right]\Phi\left(\frac{y-\mu_a}{\sigma_a}\right)}{\int_{-\infty}^{\infty}\left[\phi\left(\frac{s-\mu_0}{\sigma}\right)/\sigma\right]\Phi\left(\frac{s-\mu_a}{\sigma_a}\right)ds} \\ &+ (1-p_a)p_b\frac{\left[\phi\left(\frac{y-\mu_0}{\sigma}\right)/\sigma\right]\left[1-\Phi\left(\frac{y-\mu_b}{\sigma_b}\right)\right]}{\int_{-\infty}^{\infty}\left[\phi\left(\frac{s-\mu_0}{\sigma}\right)/\sigma\right]\left[1-\Phi\left(\frac{s-\mu_b}{\sigma_b}\right)\right]ds} \\ &+ p_ap_b\frac{\left[\phi\left(\frac{y-\mu_0}{\sigma}\right)/\sigma\right]\Phi\left(\frac{y-\mu_a}{\sigma_a}\right)\left[1-\Phi\left(\frac{y-\mu_b}{\sigma_b}\right)\right]}{\int_{-\infty}^{\infty}\left[\phi\left(\frac{s-\mu_0}{\sigma}\right)/\sigma\right]\Phi\left(\frac{s-\mu_a}{\sigma_a}\right)\left[1-\Phi\left(\frac{s-\mu_b}{\sigma_b}\right)\right]ds} \end{aligned} \quad (3)$$

Section 2: Maximum Likelihood Estimation of Experiment 2 Model Parameters

The truncation model we used to account for the results of Experiment 2 has four unknown parameters. The slope and intercept of the underlying memory standard deviation are denoted α_O and α_1, respectively. The probabilities of lower and upper truncation are p_a and p_b, respectively. We treated the uncertainty of each boundary, σ_a and σ_b, as being 50% as large as the uncertainty of stimulus values of that size. That is, $\sigma_a = 0.5(\alpha_O + \alpha_1 \times \mu_a)$ and $\sigma_b = 0.5(\alpha_O + \alpha_1 \times \mu_b)$.

(*Appendix continues*)

Our estimation procedure for the four parameters in Experiment 2 was based on maximum likelihood methods. The MLE provides an optimal fit in the sense that it is the value of each of the parameters that maximizes the likelihood of the observations. Therefore, the MLE takes into account the explicit distribution of responses predicted by the model, which is important because truncation is a nonnormal model.

Finding the MLE requires an explicit formulation of the likelihood for the observations, which are indexed by trial repetition, stimulus, and stimulus set. Denote the observed response at the *i*th trial to the *j*th stimulus in the *k*th stimulus set y_{ijk}. Denote the *j*th stimulus value in the *k*th stimulus set μ_{Ojk}. Denote the corresponding memory and boundary uncertainties σ_{jk}, σ_{ajk}, and σ_{bjk}; each is determined completely by the stimuli, the unknown parameters, and the prototypical stimulus in a category μ_{Ok}. Then the four-parameter likelihood for the truncation model used in Experiment 2 can be written in terms of the probability density given in Equation 3. It is the product of the density evaluated at each response, as follows:

$$L(\alpha_0,\alpha_1,p_a,p_b) = \prod_{\substack{i \\ 1<j<10 \\ k=1,2}} \mathrm{f}(y_{ijk}|\mu_{0_{jk}},\sigma_{jk},p_a,\sigma_{a_{jk}},p_b,\sigma_{b_{jk}}) \qquad (4)$$

where

$$\sigma_{jk} = \alpha_0 + \alpha_1\mu_{0_{jk}}$$

$$\sigma_{a_{jk}} = \alpha_0 + \alpha_1\mu_{0\cdot_k} = \sigma_{b_{jk}}.$$

We estimated parameters of the model by finding the unknown parameter values that maximize the log of the likelihood of all of the observations $\{y_{ijk}\}$ using the Newton–Raphson method, a standard numerical procedure. The standard errors of the parameter estimates are the diagonal elements of the negative inverse of the matrix of expected values of the second derivatives of the likelihood (see, e.g., Rao, 1965). We approximated these standard errors with numerical estimates of their observed values, as in the standard method of scoring.

Section 3: Truncation Is a Bayesian Procedure

Denote the inexact fine-grained value coded for a given stimulus by X, and assume it is normally distributed with mean at the true stimulus value μ_O and standard deviation σ. In Bayesian inference, information about the unknown value of μ_O prior to observing X is represented by the prior distribution $f(\mu)$. In our experiments, we assumed that this prior distribution incorporates category information equally applicable to all members of the stimulus set to which μ_O belongs. The posterior distribution $f(\mu|x, \sigma)$ represents a combination of the category information with information about the particular stimulus presented based on the fine-grained sample X. The posterior distribution of μ has the following form:

$$f(\mu|x,\sigma) = \frac{f(x|\mu,\sigma)f(\mu)}{\int_{-\infty}^{\infty} f(x|\tilde{\mu},\sigma)f(\tilde{\mu})d\tilde{\mu}}. \qquad (5)$$

As long as the category information is correct (i.e., the prior is the distribution of instances presented), the Bayes estimate given by E[μ/x, σ], the mean of the posterior distribution, is the most accurate estimate possible if only a fixed category prior distribution and an imprecise sampled X are used.

To show that truncation is a Bayesian procedure, we must show that it results in such Bayes estimates. Suppose the set of stimuli presented and the category representation thereof is defined by boundaries a and b (with $a < b$), so that it consists of all values x between a and b and that every possible value is equally likely. This is captured by the following uniform prior distribution:

$$f(\mu) = \frac{I_{a,b}(\mu)}{b-a}, \qquad (6)$$

where $I_{a,b}(x)$ is an indicator variable that takes the value 1 if $a < x < b$ and the value of 0 otherwise. Because $f(x/\mu, \sigma)$ is the normal density, it follows from Bayes' rule that the posterior density of μ is given by the following:

$$f(\mu|x,\sigma) = \frac{\left[\phi\left(\frac{\mu-x}{\sigma}\right)/\sigma\right]I_{a,b}(\mu)}{\Phi\left(\frac{b-x}{\sigma}\right) - \Phi\left(\frac{a-x}{\sigma}\right)}, \qquad (7)$$

where ϕ denotes the standard normal density and Φ denotes the standard normal cumulative distribution. A direct derivation shows that the Bayes estimate, the mean of the posterior distribution, is as follows:

$$E[\mu|x,\sigma] = x + \sigma\frac{\phi\left(\frac{a-x}{\sigma}\right) - \phi\left(\frac{b-x}{\sigma}\right)}{\Phi\left(\frac{b-x}{\sigma}\right) - \Phi\left(\frac{a-x}{\sigma}\right)}. \qquad (8)$$

The conditional posterior predictive distribution, the expected value of the Bayes estimate for a given true stimulus value $\mu_{O,}$ is the average of posterior mean values over the distribution of the sampled stimulus value X, namely as follows:

$$\int_{-\infty}^{\infty} \mathrm{E}[\mu|x,\sigma]\phi\left(\frac{x-\mu_0}{\sigma}\right)dx. \qquad (9)$$

A first-order approximation to this integral is the so-called "plug-in" estimate given by replacing x with its true mean $\mu_{O.}$ The expected value of the approximate Bayes estimate is then as follows:

$$E[\mu|x,\sigma] = \mu_0 + \sigma\frac{\phi\left(\frac{a-\mu_0}{\sigma}\right) - \phi\left(\frac{b-\mu_0}{\sigma}\right)}{\Phi\left(\frac{b-\mu_0}{\sigma}\right) - \Phi\left(\frac{a-\mu_0}{\sigma}\right)}. \qquad (10)$$

This result is identical to the mean of a truncated normal distribution with mean μ_O, namely if

$$f(y|\mu_0,\sigma) = \frac{\left[\phi\left(\frac{y-\mu_0}{\sigma}\right)/\sigma\right]I_{a,b}(y)}{\Phi\left(\frac{b-\mu_0}{\sigma}\right) - \Phi\left(\frac{a-\mu_0}{\sigma}\right)}, \qquad (11)$$

then

$$E[y|\mu_0,\sigma] = \mu_0 + \sigma \frac{\phi\left(\frac{a-\mu_0}{\sigma}\right) - \phi\left(\frac{b-\mu_0}{\sigma}\right)}{\Phi\left(\frac{b-\mu_0}{\sigma}\right) - \Phi\left(\frac{a-\mu_0}{\sigma}\right)}. \quad (12)$$

Therefore, as long as the category information is correct (i.e., the prior is the distribution of instances presented), the estimate that results from truncation is on average nearly the most accurate estimate possible with only a fixed category prior distribution and an imprecise sampled value *X*. Note that this is despite the fact that truncation introduces biased estimates. Because the mean square error of estimation is the sum of the bias squared and the variability of the estimates, the improved accuracy of a truncated estimate is due to its reduced variability.

Section 4: When Boundaries Are Inexact, Truncation Still Increases Accuracy

We reconsidered the effect of truncation on accuracy in the realistic situation in which the boundaries are uncertain and an estimate based on truncation is not used 100% of the time. When truncation boundaries are uncertain, the result is like a Bayesian procedure in which the prior distribution $f(\mu)$ imperfectly captures the distribution of instances presented. This results in a less-than-optimal estimate. Similarly, if with some probability an unadjusted sample *X* from the underlying memory is used instead of the truncated estimate, the result is intermediate in accuracy. We needed to show only that the less-than-optimal estimates produced under these circumstances are still better than the unbiased estimates from the fine-grained memory in which no category information is used, although worse than they would be if the boundaries were certain and used 100% of the time. As in the case of certain truncation occurring 100% of the time, one must lower the variability of response enough to make up for the bias such estimates introduce.

That is, to show that truncation still improves accuracy under realistic conditions, we needed to verify the second half of the following inequality:

$$E_{f(y|\mu,\sigma)}[(y-\mu)^2] < E_{f(y|\mu,\sigma,p_a,\sigma_a,p_b,\sigma_b)}\,[(y-\mu)^2] < \sigma^2. \quad (13)$$

Here the minimum expectation denotes the standard error of estimates based on the approximate Bayes estimate yielded by certain truncation at the stimulus extrema *a* and *b*, whereas the intermediate expectation is the standard error of estimates based on uncertain truncation with probabilities p_a and p_b and imprecise boundaries with standard deviations σ_a and σ_b.

We directly verified that the second inequality holds in the context of Experiment 2—so that even suboptimal truncation improves the accuracy of estimates—using a Monte Carlo simulation. We focused in particular on the case in which lower truncation occurs 100% of the time, whereas upper truncation does not, and the boundary uncertainties are assumed to be a fraction of the underlying memory uncertainty, as in the model of Experiment 2. We sampled 10,000 stimulus values, half uniformly from each of the two ranges of stimuli (from 5 cm^2 to 50 cm^2 and from 55 cm^2 to 100 cm^2). Samples from the underlying memory were drawn with standard deviations set in linear relation to the stimuli with the parameters estimated in Experiment 2 and given in Table 3. Five levels of boundary uncertainty were set, from 0% (uncertain boundaries) to 100% of the standard deviation of the underlying memory at the prototypical stimulus in each stimulus set. The probability of lower truncation was set to 100%, whereas the probability of upper truncation was allowed to vary between 0% and 90% in 10% increments and between 90% and 100% in 1% increments.

Table 3 shows (as percentages) the ratio of the mean square error of 10,000 truncated estimates to the mean square error of a separate set of 10,000 unadjusted samples from the underlying memory, based on identical uniformly distributed stimuli. A lower percentage reflects a smaller error and greater improvement on unadjusted sampling. When upper truncation is 100% probable, as at the top of the table, the adjusted estimates are best for any given level of boundary uncertainty. Moving down the table, as truncation occurs with a lesser probability, the percentages increase so that adjusted estimates improve less on unadjusted estimates. Truncation leads to more inaccurate estimates only where entries in the table exceed 100%, so that the adjusted estimate has a larger standard error than unadjusted samples from the underlying memory. Note that for the vast majority of probabilities of upper truncation and boundary uncertainties, the adjusted estimate is better than the unadjusted sample from the underlying memory. Truncation only worsens the accuracy of estimation for the psychologically implausible case that boundaries are as uncertain as the fine-grained memory, as in the last column of the table.

The setting for upper truncation probability and boundary uncertainty that were most similar to the fitted model in Experiment 2 is shown in bold. We can then predict that if estimates were derived from an uncertain, unbiased memory and were truncated, their accuracy increased by more than 1.5 times, as their mean square error decreased by more than a third. Therefore, the model predicts that participants increased their accuracy by using truncation, even though they did not use upper truncation to as great an extent as lower truncation and despite the apparent uncertainty of the boundaries. Furthermore, although the uncertainty of truncation boundaries was set in an ad hoc fashion in accounting for the results of Experiment 2, it appears that any reasonable setting (along with evidence that truncation is occurring with a high probability) would lead us to the same conclusion.

Received July 1, 2003
Revision received November 15, 2006
Accepted November 20, 2006 ■