Category effects on stimulus estimation: Shifting and skewed frequency distributions

SEAN DUFFY

Rutgers University, Camden, New Jersey

JANELLEN HUTTENLOCHER

University of Chicago, Chicago, Illinois

LARRY V. HEDGES

Northwestern University, Evanston, Illinois

AND

L. ELIZABETH CRAWFORD

University of Richmond, Richmond, Virginia

The category adjustment model (CAM) proposes that estimates of inexactly remembered stimuli are adjusted toward the central value of the category of which the stimuli are members. Adjusting estimates toward the average value of all category instances, properly weighted for memory uncertainty, maximizes the average accuracy of estimates. Thus far, the CAM has been tested only with symmetrical category distributions in which the central stimulus value is also the mean. We report two experiments using asymmetric (skewed) distributions in which there is more than one possible central value: one where the frequency distribution shifts over the course of time, and the other where the frequency distribution is skewed. In both cases, we find that people adjust estimates toward the category's running mean, which is consistent with the CAM but not with alternative explanations for the adjustment of stimuli toward a category's central value.

This article explores a well-known finding in the memory literature: that estimates of categorized stimuli are often remembered as being more typical members of their categories than they actually are. Known variously as the *central tendency bias* or *schema effect*, this phenomenon has been described by several psychologists as perceptual or memory distortions (Bartlett, 1932; Estes, 1997; Hollingworth, 1910; Poulton, 1979). Alternatively, Huttenlocher and colleagues (e.g., Crawford, Huttenlocher, & Engbretson, 2000; Huttenlocher, Hedges, & Vevea, 2000) have proposed a rational basis for these effects. They argued that this bias arises from an adaptive Bayesian process that improves accuracy in estimation. The category adjustment model (CAM) proposes that stimuli are encoded at two levels of detail: as members of a category, and as fine-grain values. In reconstructing stimuli, people combine information from both category and fine-grain levels of detail. This combination results in estimates adjusted toward the central region of their categories. This adjustment reduces the mean square error of estimates at any given stimulus value enough to more than compensate for the bias introduced into individual estimates (Huttenlocher et al., 2000, p. 240).

In their model, a category is a bounded range of stimulus values that vary along a stimulus dimension, such as

size, weight, or intelligence. Memory for a stimulus is a fine-grain value along a dimension, such as a specific person's height. If the remembered value is inexact, the model proposes that the estimate (R) of the stimulus is a weighted combination of a category's central value (ρ) and the inexact fine-grain memory (M) for a particular stimulus. The weight λ given to the fine-grain and category levels varies as a function of the dispersion of the category (σ_ρ^2) and the degree of inexactness surrounding the fine-grain memory (σ_M^2) and is derived from Bayes's theorem. To illustrate the Bayesian principle underlying the model, consider the example in Figure 1. The figure depicts a category with a normal frequency distribution of instances that varies along a continuous dimension. If a fine-grain memory for a stimulus falls at value M , and there is uncertainty surrounding the stimulus's true value, it is more likely that the stimulus's true value is in Direction A (toward the direction where the majority of instances fall) than in Direction B (where there are fewer instances). The combination of information about the prior distribution with the present distribution of inexactness surrounding the true value for M results in biased estimates that are more likely to fall in Region A than in Region B.

S. Duffy, seduffy@camden.rutgers.edu

Figure 1. The category adjustment model (CAM).

Huttenlocher et al. (2000) tested the CAM with an experimental task in which participants learned an inductive category by observing and reproducing a series of stimuli, such as lines that varied in length. On each trial, a target line briefly appeared; participants estimated its length, after a delay, by adjusting a response line to be the same length. Participants were presented with a series of these lines that varied in length, drawn in random order from a symmetric distribution of lines. After only a small number of trials, participants began adjusting responses toward the central value of the distribution of lines, overestimating short lines and underestimating long lines. Moreover, the extent of the adjustment varied as a function of the frequency distribution of the observed set, with stronger bias where the distribution was peaked rather than uniformly distributed.

Although Huttenlocher et al. (2000) suggested that people adjust responses toward the category's central value, which is presumed to be the running mean of all instances, there is an alternative explanation for this observed bias: that overall category information is not used in reproducing estimates but arises from a memory distortion process in which the fine-grain memory for a stimulus on any given trial is "blended" with the fine-grain memory of the immediately preceding stimulus or some set of recent stimuli. Averaged across all trials, such distortions would result in a central tendency bias. Consider why by returning to Figure 1. For any stimulus value other than the centermost value, the immediately preceding stimulus has a higher probability of having been in the direction of the central value or beyond it (Region A) than in the sparsely populated tail of the distribution (Region B). Thus, if people produce estimates

that are a blend between the magnitude of a current and a previous stimulus, on average they will adjust larger stimuli downward and smaller stimuli upward, resulting in a central tendency bias.

The claim that one or a few prior instances are used in estimation comes from several studies in both the psychophysics and memory literatures that explore sequential effects in judgment. In a serial reproduction task similar to those used by Huttenlocher et al. (2000), Sailor and Antoine (2005) found that sequential estimates of squares that varied in size were influenced by the magnitude of the immediately preceding square. In psychophysical rating tasks in which people assign numerical values to squares of varying sizes (DeCarlo & Cross, 1990; Petzold & Haubensak, 2004), researchers typically observe an effect of the most recent stimulus as well as the adaptation level (Helson, 1964) of the whole series. Because of these findings, it is necessary to test whether such memory blending explains the effects reported by Huttenlocher et al.

A useful approach for determining whether people use the running average of all stimuli or blend estimates of a target stimulus with a subset of recent stimuli is employing distributions that either shift over time or have a skewed frequency distribution. Such distributions are useful in that they allow us to separate the influence of the running average of all stimuli from the influence of more recent instances. In Experiment 1, we examined whether people use an entire distribution or only recent instances when estimating stimuli by shifting a category's distribution during the experiment. Participants initially estimated stimuli embedded within a frequency distribution exhibiting either positive or negative skew. In the middle of the experiment and without indication, they begin estimating

stimuli from a distribution with the opposite skew (i.e., those who start with a positive skew distribution begin estimating stimuli from a distribution exhibiting negative skew). This design allowed us to test whether people adjust responses toward the running mean of all instances or use only the most recent or the average of a subset of recent stimuli. If people blended their inexact memories of a stimulus on any given trial with the stimulus on one or a few previous trials, their estimates in the second half of the experiment would be adjusted toward the mean of the distribution used in the second half of the experiment. Alternatively, if people used the running average of the entire distribution, the estimates for the second half of the experiment should be affected by the distribution in the first half. Participants would bias responses in the second half toward a region along the stimulus dimension that is intermediate between the mean of the distributions used in the first and second halves.

Having found evidence that people bias responses toward the running average of previous instances, in Experiment 2, we explored what statistical measure of the category center people use as the central category value. Prior studies of the CAM assume that people use the mean of the distribution of prior instances (e.g., Duffy, Huttenlocher, & Crawford, 2006; Duffy & Kitayama, 2007; Huttenlocher et al., 2000). These studies used symmetric stimulus distributions, in which the various measures of centrality converged on the same category value in the middle of the distribution; therefore, these studies could not determine what measure of central tendency people used. In Experiment 2, participants estimated stimuli embedded within positively or negatively skewed frequency distributions. With such skewed distributions, different statistical measures of centrality fall at different points along the distribution of stimuli, enabling us to determine what statistical representation of centrality best represents the value participants used to establish the center of the category.

EXPERIMENT 1

The purpose of Experiment 1 was to determine whether participants adjusted responses toward the mean of all stimuli presented or to some other point such as the mean of a small number of recent stimuli from a recent subset (e.g., the last 1, 2, 3, . . . , 10 stimuli).

Method

Participants. Twenty-five undergraduates participated in Experiment 1. There were 13 males and 12 females.

Design and Procedure. Participants were told to reproduce the length of lines from memory. Stimuli were presented on a 19-in. CRT monitor (70 ppi resolution). On each trial, participants were shown a randomly chosen line from the distribution for 1 sec. After a 1-sec delay, an adjustable line appeared that participants could make longer or shorter by using the left and right mouse buttons. The initial length of this adjustable line varied between participants, starting at either 40 or 400 pixels (i.e., 32 pixels beyond the range of the stimulus distribution). This manipulation was to ensure that there was no overall over- or underestimation of the lines due to the starting length of the reproduction line; the starting length of the

Table 1
Experiment 1 Stimulus Distributions

Stimulus Number	Stimulus Size	Positive Skew	Negative Skew	Total
1	80	9	1	10
2	96	9	1	10
3	112	8	2	10
4	128	8	2	10
5	144	7	3	10
6	160	7	3	10
7	176	6	4	10
8	192	6	4	10
9	208	5	5	10
10	224	5	5	10
11	240	5	5	10
12	256	4	6	10
13	272	4	6	10
14	288	3	7	10
15	304	3	7	10
16	320	2	8	10
17	336	2	8	10
18	352	1	9	10
19	368	1	9	10

response line had no significant influence on the resulting estimates ($F < 1$). After participants were satisfied with the length of the reproduction line, they pressed "Enter," beginning the next trial after a 1-sec pause.

The experiment ended once participants reproduced all 190 lines. Participants were randomly assigned to one of two counterbalanced orders. For the first 95 of the 190 trials, participants estimated stimuli embedded within a positively or negatively skewed distribution. Without interruption, at the 96th trial, the distribution shifted, so that participants who had initially estimated stimuli from a positively skewed distribution immediately began estimating stimuli from a negatively skewed distribution, and vice versa. In each of the two distributions, there were 19 distinct stimulus values ranging from 80 to 368 pixels in length in 16-pixel increments. The frequency distributions are presented in Table 1. By combining these two distributions, participants ultimately estimated an equal number of stimuli (10) from each of the 19 stimulus values.

Results and Discussion

For each trial, bias was calculated as the difference between the participant's estimated length and the actual stimulus length. Positive bias indicates stimulus overestimation; negative bias, stimulus underestimation. The left panels of Figure 2 depict the frequency distributions of the two halves of the experiment, and the right panels depict the average bias for each stimulus value for both halves of the experiment. Note that, in both conditions for the first 95 trials, the bias curve crosses the x -axis near the mean of the distribution used for the first 95 trials. However, the bias curves for Trials 96–190 cross the x -axis at a point intermediate between the means of the distribution used for the first 95 trials and Trials 96–190. This suggests that category information acquired during the initial 95 trials influenced estimates during the second half of the experiment once the distribution shifted, consistent with a model in which the running average of all trials is used as the category's central value.

To quantitatively test whether people adjust stimuli toward a few preceding stimuli or toward the mean of all stimuli, we carried out a regression analysis on each indi-

Figure 2. Results from Experiment 1. The two histograms depict the frequency distributions for the two halves of the experiment. The panels in the far right are scatterplots of bias by condition. P.S., positive skew distribution; N.S., negative skew distribution.

vidual participant's estimates on three predictor variables (with no intercept): the actual stimulus value, the mean of all previous stimulus values, and the mean of the preceding (1–20) stimulus values. That is, we fitted the model $R = \lambda_1(\text{stimulus}) + \lambda_2(\text{mean}) + \lambda_3(\text{preceding})$. Note that the crucial assumption of the CAM model described above is that $\lambda_3 = 0$. Therefore, we tested this assumption by testing the hypothesis that $\lambda_3 = 0$.

For all regressions including the previous 1–20 stimuli, the pattern of results is the same in all analyses. The current stimulus has a large and highly statistically significant impact on the responses, receiving about 80% of the weight (e.g., $\lambda_1 = .8$, in each case). The mean of all stimuli has a smaller but statistically significant impact in all cases, receiving 13% to 16% of the weight (i.e., $\lambda_2 = .13$ to $.16$). The impact of the preceding 1 to 20 stimuli is much smaller and statistically insignificant ($p > .1$) in every analysis. These results are consistent with the CAM model's prediction that recent stimuli will receive no more weight than they receive as part of the overall average. They are inconsistent with accounts arguing that the central tendency bias is a distortion caused by the immediately preceding stimuli (e.g., Choplin & Hummel, 2002).

EXPERIMENT 2

Experiment 2 aimed to expand on the findings of Experiment 1 by examining what measure of central tendency people use as the category center for asymmetric distributions. The CAM suggests that the mean of all prior

instances serves as the central value of the category. However, people may use other measures of central tendency, such as the median, mode, or center of the range. With symmetric stimulus distributions used in most prior studies, these measures of central tendency are identical. However, with asymmetric distributions, the statistical measures of central tendency fall at different points along the distribution. Experiment 2 expanded on Experiment 1: It tested more directly whether people adjust responses toward the running mean of all instances by using skewed stimulus frequency distributions that did not shift over time.

Method

Participants. Thirty-six undergraduate students participated in Experiment 2. The sample included 20 males and 16 females.

Design and Procedure. In each of the distributions, there were 19 distinct stimulus values, ranging from 80 to 368 pixels in length in 16-pixel increments. In the positive skew condition, there were 19 instances of Stimulus 1, 18 instances of Stimulus 2, and so on, down to 1 instance of Stimulus 19. In the negative skew condition, there was 1 instance of Stimulus 1, 2 of Stimulus 2, and so on, up to 19 instances of Stimulus 19. In the uniform condition, there were 10 instances of all 19 stimulus values. These are depicted in the top half of Figure 2.

Results and Discussion

Bias was calculated as in Experiment 1. Average biases across all participants for each stimulus value for the three conditions are shown at the bottom of Figure 3. The dots represent the average bias for each of the 19 stimulus values for all participants for the three conditions.

Figure 3. Results for Experiment 2. In the top panels are the frequency distributions, and arrows indicate the value for the arithmetic mean of the distribution. The bottom panels contain scatterplots of bias for each condition.

The CAM model implies that stimuli are adjusted toward the mean of all category instances. Therefore, the point toward which participants adjust estimates can be operationalized as the point along the *x*-axis, where bias equals zero. To locate this point, we regressed bias on actual stimulus size and estimated the point at which the bias was zero, following regression procedures outlined by Stuart, Ord, and Arnold (1999, p. 125) for determining at what point along the *x*-axis bias equals zero and the 95% confidence intervals surrounding that value.

As can be seen in the bottom of Figure 3, the bias curves cross the *x*-axis near the mean of all stimulus values. The estimated point of zero bias (and its 95% confidence interval) is presented for all three conditions in Table 2. In all three conditions, the 95% confidence interval for the estimated point of zero bias included the arithmetic mean of the distribution. The critical finding is that for neither of the skewed distributions did the confidence interval for the estimated point of zero bias include the median, mode, or center of the range of stimuli. This implies that participants adjusted their estimates toward the arithmetic mean of the stimuli, rather than toward other statistical measures of centrality.

Additional Analyses of Experiment 1 in Relation to Experiment 2

The shape of stimulus distributions in Experiment 1 changed halfway through the experiment. This provided an opportunity to further test the hypothesis investigated in Experiment 2; that is, we could test whether the participants adjusted their responses toward the mean or some

other statistical summary of the distribution of stimuli in each half of the experiment. To do so, we examined the estimated point of zero bias in the first and second halves of the experiment separately, using the procedure outlined by Stuart et al. (1999). The resulting estimates are provided in Table 3. The point of zero bias for each stimulus value for Trials 1–95 and for Trials 96–190 was computed independently. For the negative-to-positive skew condi-

Table 2
Estimates for the Point of Zero Bias for the Three Distributions in Experiment 2

Distribution	Actual Distribution			Participant Responses		
	<i>M</i>	Median	Mode	Zero Bias	95% CI	
Positive skew	176	160	80	178.77	175.89	181.64
Uniform	224	224	224	223.14	221.42	224.86
Negative skew	272	288	368	269.63	265.23	274.04

Table 3
Estimates for the Point of Zero Bias for the First and Second Halves of Experiment 1

Condition	Actual <i>M</i>	Zero Bias	95% CI	
			Lower	Upper
Trials 1–95				
Pos–neg	180.21	178.78	177.00	180.53
Neg–pos	267.79	267.17	265.64	268.72
Trials 96–190				
Pos–neg	267.79	226.09	224.68	227.49
Neg–pos	180.21	218.52	216.33	220.72

tion, the point of zero bias for the first 95 trials fell along the x -axis at 267 pixels, exactly at the mean for the negative skew distribution. For the positive-to-negative skew condition, the point of zero bias for the first half fell along the x -axis at 179 pixels, very near to the true mean of the distribution (180.21). In both cases, the 95% confidence interval for the point of zero bias included the true mean of the distribution of instances. Importantly, the 95% confidence intervals do not include either the median stimulus value (160 for the positive skew condition, 288 for the negative skew condition) or the mode (88 for the positive skew condition, 368 for the negative skew condition). Clearly, people adjust responses toward the mean of the entire prior distribution of stimuli.

For the second half of the experiment, participants estimated stimuli embedded in a distribution exhibiting the opposite skew from that in the first half of the experiment. For the positive-to-negative condition, the point of zero bias is 226.09, far from the actual mean of this half of the distribution (267.79). For the negative-to-positive condition, the point of zero bias is 218.52, also far from the true mean of the distribution (180.21). For neither of these two conditions do the 95% confidence intervals include the mean of that half of the distribution. However, for both conditions, the point of zero bias for the second half of the experiment is close to the average of the entire distribution (224), consistent with the notion that people adjust estimates toward the mean of the entire set of instances.

Importantly, these results provide direct evidence that the central tendency bias in stimulus estimation cannot be explained as a memory blend between the magnitude of a target stimulus and a small set of stimuli immediately preceding it. In the latter case, the point of zero bias for both halves of the experiment should have fallen at the respective means of the two distributions, since information about stimuli experienced in the first half of the experiment would not influence estimates in the second half. The fact that the point of zero bias for the second half of both conditions of the experiment fell near the mean of the entire distribution suggests that people continued to update their initial representation throughout the experiment, adjusting responses toward the average of all stimuli encountered.

GENERAL DISCUSSION

The two experiments reported here support a model of stimulus estimation we presented earlier, the CAM. The results indicate that people adjust estimates of stimuli toward the mean of the entire category distribution, even when the category's frequency distribution shifts over time (Experiment 1). They do not adjust responses toward other possible measures of centrality such as the median, the mode, or the center of the stimulus range, even in conditions where the category's frequency distribution is asymmetric (Experiment 2).

The results of Experiment 1 provide evidence that the central tendency bias in estimation cannot be explained by a process in which the fine-grain memory of a cur-

rent stimulus is blended with, or distorted by, immediately preceding stimuli. Rather, in Experiment 1, participants appeared to continually update their representation of a distribution of stimuli by integrating information about new stimuli into their category representation and adjusting estimates toward the running average of the distribution. Note, however, that this finding does not necessarily indicate that there is no influence of the prior stimulus, only that such an influence is far smaller than is the influence of the entire distribution. Importantly, Sailor and Antoine (2005), as well as DeCarlo and Cross (1990), also obtained evidence that information about the sequence as a whole (i.e., adaptation level) affected judgments in the tasks they used. The magnitude of the immediately preceding stimulus also affected estimates, but the influence of the immediately preceding stimulus was minor, relative to the influence of the entire distribution; so there is no strong evidence for an explanation of the central tendency bias as a memory distortion caused by a subset of immediately preceding stimuli.

The results of Experiment 2 allowed us to differentiate between possible values that might have been used as the category's central value. Prior studies, using only symmetrically distributed categories, did not allow us to determine which statistical representation of central tendency people used. Our findings in Experiment 2 with skewed distributions confirm the results obtained in Experiment 1: People adjust estimates toward the mean of the category distribution. This is consistent with the Bayesian assumptions underlying the CAM. The finding also suggests that estimates reflect sensitivity to the shape of the frequency distribution of stimuli, consistent with studies in different stimulus domains (e.g., Flannagan, Fried, & Holyoak, 1986; Hasher & Zacks, 1979).

The CAM was applied earlier to cases in which a category's frequency distribution remained the same over time. What we have found here is that people use a single overall category to adjust estimates, even in situations where the frequency distribution shifts. Such shifts may be common, due to natural or artificial processes; for instance, the height of a class of children increases over the academic year, or houses become larger as one drives from a city to its suburbs. In such situations, it may be more adaptive to ignore information learned early in the process of category induction and rely more on information about more recent stimuli. Such "criterion shifts" in distributions may be an important direction for future research on estimation (Brown & Steyvers, 2005; Duffy & Crawford, 2008).

Furthermore, the findings of the present study concern estimates of simple perceptual stimuli that vary along a single continuous dimension. Different types of stimuli exhibit unique psychophysical properties that may influence the category value subjectively experienced as the distribution's center. We used lines that varied in length because they have a power law exponent of 1, so that objective and subjective magnitudes are similar (Stevens, 1957). Future work on the estimation may explore how psychophysical properties of various stimuli affect estimation processes.

AUTHOR NOTE

The authors thank Michael Lee and two anonymous reviewers for their helpful comments on an earlier version of the manuscript. Address correspondence to S. Duffy, Rutgers University, 343 Armitage Hall, 411 N. 5th Street, Camden, NJ 08102 (e-mail: seduffy@camden.rutgers.edu).

REFERENCES

- BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- BROWN, S. D., & STEYVERS, M. (2005). The dynamics of experimentally induced criterion shifts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 587-599. doi:10.1037/0278-7393.31.4.587
- CHOPLIN, J. M., & HUMMEL, J. E. (2002). Magnitude comparisons distort mental representations of magnitude. *Journal of Experimental Psychology: General*, **131**, 270-286. doi:10.1037/0096-3445.131.2.270
- CRAWFORD, L. E., HUTTENLOCHER, J., & ENGBRETSON, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, **11**, 280-281. doi:10.1111/1467-9280.00256
- DECARLO, L. T., & CROSS, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*, **119**, 375-396. doi:10.1037/0096-3445.119.4.375
- DUFFY, S., & CRAWFORD, L. E. (2008). Primacy or recency effects in the formation of inductive categories. *Memory & Cognition*, **36**, 567-577. doi:10.3758/MC.36.3.567
- DUFFY, S., HUTTENLOCHER, J., & CRAWFORD, L. E. (2006). Children use categories to maximize accuracy in estimation. *Developmental Science*, **9**, 597-603. doi:10.1111/j.1467-7687.2006.00538.x
- DUFFY, S., & KITAYAMA, S. (2007). Mnemonic context effect in two cultures: Attention to memory representations? *Cognitive Science*, **31**, 1009-1020.
- ESTES, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review*, **104**, 148-169. doi:10.1037/0033-295X.104.1.148
- FLANNAGAN, M. J., FRIED, L. S., & HOLYOAK, K. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 241-256. doi:10.1037/0278-7393.12.2.241
- HASHER, L., & ZACKS, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, **108**, 356-388. doi:10.1037/0096-3445.108.3.356
- HELSON, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- HOLLINGWORTH, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology, & Scientific Methods*, **7**, 461-469.
- HUTTENLOCHER, J., HEDGES, L. V., & VEVEA, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, **129**, 220-241. doi:10.1037/0096-3445.129.2.220
- PETZOLD, P., & HAUBENSAK, G. (2004). The influence of category membership of stimuli on sequential effects in magnitude estimation. *Perception & Psychophysics*, **66**, 655-678.
- POULTON, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, **86**, 777-803. doi:10.1037/0033-2909.86.4.777
- SAILOR, K. M., & ANTOINE, M. (2005). Is memory for stimulus magnitude Bayesian? *Memory & Cognition*, **33**, 840-851.
- STEVENS, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153-181. doi:10.1037/h0046162
- STUART, A., ORD, J. K., & ARNOLD, S. (1999). *Kendall's advanced theory of statistics: Classical inference and the linear model*. London: Hodder Arnold.

(Manuscript received September 14, 2007;
revision accepted for publication November 9, 2009.)