THE EFFECT OF CATEGORY ADJUSTMENT MODEL (CAM) ON THE STIMULUS ESTIMATION: HOW ASPECTS OF THE RESPONSE STIMULUS AFFECTS TASK PERFORMANCE

by

SUYEON NOH

A thesis submitted to the

Graduate School-Camden

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of Master of Art

Graduate Program in Psychology

written under the direction of

Sean Duffy

and approved by

Sean Duffy

Robrecht van der Wel

Sarah Allred

Camden, New Jersey (May, 2013)

ABSTRACT OF THE THESIS

The Effect of Category Adjustment Model (CAM) on the Stimulus Estimation:

How aspects of the response stimulus affects task performance

BY SUYEON NOH

Thesis Director:

Sean Duffy

Studies on the category adjustment often use sequential stimulus estimation tasks are in which people see a target stimulus then must reproduce its size using an adjustable response stimulus. Two empirical questions remain regarding methodological considerations in such tasks. In this study, the starting length and velocity of response stimulus as well as the time interval between target presentation and response are investigated. We found that velocity and response delay affect estimates but starting size does not. (77 words)

The Effect of Category Adjustment Model (CAM) on the Stimulus Estimation: How aspects of the response stimulus affects task performance

Introduction

It is important to remember and recall information accurately in daily life. We are often required to report a stimulus' value along some continuous dimension, such as weight, number, length, or hue. For example, in cooking, one might need to recall the amount of celery used in a soup, the number of apples used in baking a pie, the correct length to cut carrots, or the proper shade of a particular beer. However, any given memory we have, there is certain to be inexactness surrounding it. This is in part because our perceptual and memory systems are not perfect, and this natural imperfection of sensory perception or cognitive processes could cause inexactness to enter the representation. In such cases of inexactness, people often rely upon information stored in categories to fill in for the information that is missing. This is a rational strategy for minds to use because categories contain rich information that can be used to fill in those aspects of an individual memory that are impoverished or entirely missing.

The category adjustment model (Huttenlocher, Hedges, & Vevea, 2000) is a Bayesian model of memory that posits that an estimate of an inexactly encoded stimulus is a combination of the fine-grain but imprecise recollection of the stimulus' value on a dimension with category-level information about the typical value that category members tend to exhibit along that value. For example, imagine you meet a new colleague at a conference whom you have never encountered previously. You remember a variety of things along a number of dimensions about this colleague, such as his height, weight, or skin tone. Later, you may be asked to recall his value along one of these dimensions, such as his height. But your memory of his height may be foggy; it is unlikely that you measured him or specifically asked him for his height. Now imagine his actual height is 72 inches, and the average male is 68 inches. You will likely remember his height to be intermediate between his actual height and the category prototype, for instance, as 70 inches. Figure 1 presents a visual depiction of the model.

There have been a number of empirical studies that have tested the robustness of the category adjustment model (Huttenlocher et al., 2000; Crawford, Huttenlocher, & Engebretson, 2000; Duffy, Huttenlocher, & Crawford, 2010). Most use a task known as the sequential stimulus estimation task. In this paradigm, participants view and reproduce a series of lines that vary in length. Over time, they develop a transient category of line lengths and begin adjusting responses toward the prototypical stimulus value, shortening long line lengths and lengthening short line lengths. Yet across these experiments, conducted in a variety of laboratories, there are some inconsistencies in the experimental methodologies used that might affect the comparison of results. The aim of the present study is to determine whether two aspects of the design of these experiments affect results: the starting length of response lines and the length of delay between target presentation and response.

Background of the Category Adjustment Model

The origin of the model used in the present study has its origin in helping to explain a series of memory distortions in time and space that later became adopted for the kinds of categories we learn through inductive experience. The category adjustment model was first formally defined in a paper on memory for temporal events to explain what was known as "forward telescoping" (Huttenlocher, Hedges, & Prohaska, 1988, 1992; Huttenlocher, Hedges, & Bradburn, 1990). Forward telescoping is a phenomenon in which people tend to remember discrete events as having happened more recently than they actually occurred. So in February, one will remember having gone to a party in August that actually was held in July. This bias in memory was explained due to the fact that the present moment serves as a strong category boundary while the past has very few salient boundaries. Given the inexactness surrounding the true memory for the date of an event, the fact that it is memorable at all would suggest that it occurred more recently than it actually occurred relative to days in which no memorable events occurred. Huttenlocher et al. used a clever task to explore these temporal biases by exploiting the trimester system used at the University of Chicago and the popularity of a campus film organization known as Documentary Films, which featured films at a steep discount ($2 in 1990) on a daily basis that most Chicago students took advantage of (Duffy, personal communication, 2013).

In most studies of forward telescoping, participants are asked about events in the past in which the past was relatively unbounded. However, in the lives of college students, their temporality is bounded by the beginning and ends of semesters. In their study, Huttenlocher et al. took advantage of this aspect of college life, as well as the popularity of the documentary film group, by asking students in a phone interview which films they remembered seeing in the past year and when they remembered seeing them. They found that close to semester boundaries, people were accurate at reporting their memories for when they saw a given film, yet toward the middle half of semesters, people tended to bias their reports toward the mid-point of the semester (see Figure 2). Hence, semesters served as temporal categories with discrete boundaries yet inexact central regions, and so the effect of forward telescoping that was found previously showed a different pattern for individuals whose lives were temporally bounded into categories.

Later, Huttenlocher and her colleagues extended the category adjustment model to spatial categories. Huttenlocher, Hedges, and Duncan (1991) explain a series of biases in reporting memories of location within a circle. In their studies, they found that people tended to bias their memories of the location of stimuli toward the center of the Cartesian quadrants of a circle but not at locations near the X or Y axes that subdivide the circle (see Figure 3). This was explained by the fact that points close to the X or Y axis are able to be coded more exactly than those that are near the diagonals. So when coding location, people code location as an inexact fine grain memory as well as a category (the quadrant in which the dot appeared). People combine those two levels of information on estimation resulting in responses that are biased toward the center of the quadrants for points that are not near the X or Y axis. Those near the quadrant boundaries, since they are coded more precisely, are biased less. This was explored in greater depth in a series of follow up studies (Engebretson & Huttenlocher, 1996; Huttenlocher & Hedges, 1994; Sandberg, Huttenlocher, & Newcombe, 1996; Tversky & Schiano, 1997).

Most germane to the present study, Huttenlocher and her colleagues extended the model from temporal and spatial categories to explain distortions in memory found in estimating stimuli from a category whose members vary along some dimension. There is a well-known finding in the memory and psychophysical literature that people tend to remember stimuli as being more typical of the category of which they are members than they actually are. Hollingworth (1910) first described this central tendency bias in stimulus estimation, in which people tend to exhibit a 'regression to the mean' in responding to stimuli that vary along a continuum, lengthening short stimuli and shortening long stimuli. Over the course of the 20th century, a variety of proposals were offered to explain such biases, most of them attributing this phenomenon to a memory or perceptual distortion (Goldstron, 1994; Parducci, 1965) or attributed the bias to adaptation level theory (Helson, 1964). For instance, Petzold & Haubensak (2004) and Sailor and Antoine (2005) explain the central tendency bias in estimation as influence exerted by an immediately preceding stimulus that interferes with the memory of the current stimulus. So for many years, such distortions in memory were explained as flawed processing.

Rather than conceptualizing the central tendency bias as a distortion, Huttenlocher and her colleagues (Huttenlocher et al., 2000; Crawford and Huttenlocher, 1996; Duffy, Huttenlocher, Hedges, & Crawford, 2008; Duffy & Crawford, 2010) have explained the central tendency bias as an adaptive mechanism that over time increases the accuracy of stimulus estimation. In the category adjustment model, category is defined as clusters of relevant stimuli dimensions that consist of a cognitive structure, and stimulus is considered as a value along the relevant set of dimensions (Huttenlocher et al., 2000). Categories typically have lower and upper boundaries that represent the smallest or largest possible stimulus size (i.e., the smallest and tallest building within a city), and the center of the category being the most typical member of the category. For most categories, this is the average value of the category (see Duffy, et al., 2008 for a discussion of categories that do not exhibit symmetric frequency distributions). Stimuli are encoded as a fine-grain memory along the relevant stimulus dimension (x is *about* 35 feet tall), and as a member of a category (x is a tree).

Upon recall, information from both levels is combined to create an estimate of the stimulus (x is about 35 feet tall; trees are on average about 25 feet tall, hence to be safe I will estimate x as 30 feet). To illustrate, consider figure 1. It depicts a category of instances that vary along a continuous dimension, such as tree heights. The category average (p) is the typical height of trees. Imagine we need to recall tree x, which was actually 35 feet (M), but since it was encoded imprecisely, has a distribution of inexactness surrounding its true value. Given this inexactness, it is more rational to select a value to the left of M since the vast majority of category members fall in that direction as opposed to the right, where there are very few.

The model is Bayesian in that it uses a prior distribution (a category) to adjust the inexactness of a present distribution (the error surrounding the fine grain memory). It is a precisely specified yet elegantly simple model, described by the following equation:

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

Where R is the stimulus response, M represents the average of a set of values in memory for the stimulus, p is the central value of the category, and λ is a weighting parameter that ranges from 1 to 0 (Huttenlocher et al., 2000; Sailor et al., 2005). In the model, λ is a function of the variability of the category itself as well as the variability of the error surrounding the inexact memory. It too is precisely specified, and is defined as:

$$\lambda = \sigma^2_\rho / (\sigma^2_\rho + \sigma^2_M),$$

Where σ^2_ρ represents the variability of the category and σ^2_M represents the degree of inexactness in the memory. So when σ^2_M is small when the fine grain memory is precise, λ is close to 1, and response is closer to the fine grain memory. Yet when σ^2_M approaches infinity, when nothing is known about the stimulus' size, λ is close to 0, and only the category average is used.

The model has an additional level concerning the kurtosis or shape of the distribution. When categories have many instances at the center of a category yet few at the edges (i.e., a leptokurtic distribution), the extent to which people bias responses toward the center of the category decreases relative to a category in which the members vary widely (i.e., a platykurtic distribution). This is possibly because people intuitively assume that extreme members of a category must be members of some adjacent category, so biasing toward the average of the wrong category would lead to extreme inaccuracy in estimation. Alternatively, extreme category instances might garner more attentional resources because they are unusual category exemplars, and hence are encoded more precisely. Thus, the bias curve takes on a curvilinear shape, resembling a sideways S. This is a similar phenomenon to what happens with the encoding of location in a circle near the X and Y axis; their proximity to these oblique axes of symmetry allows them to be encoded with greater precision. This issue is explored in greater depth in an article by Huttenlocher, Hedges, Lourenco, Crawford, and Corrigan (2007) but is somewhat unrelated to the present investigation.

An important aspect of the model is that bias arises at the point of reconstruction as opposed to at the time of encoding, as has been suggested by Goldstone (1994). This was investigated in depth in a research article by Crawford and her colleagues (2000) which addressed when the central-tendency bias occurs. The focus of this investigation was under two different conditions. Capitalizing on the Mueller-Lyer illusion, the study compared the effect of a perceptual illusion with the central-tendency bias through using three different shapes of line –no arrowheads, arrowheads pointing in, and arrowheads pointing out- under the two conditions – perceptual learning account and stimulus reconstruction account (Crawford et al., 2000). It was hypothesized that if the central-tendency bias occurs only at encoding stage, it should not affect any different stages of memory such as whether or not the stimulus is held in memory before being adjusted. In contrast, if the central-tendency bias arises after encoding the stimulus, it should increase with added memory strategy (Crawford et al., 2000). The participants consisted of 70 undergraduate students. They were asked to reproduce the lines shown with two conditions: whether the stimulus remained in-view condition while estimating it, or whether it disappeared and needed to be reconstructed based on the memory (memory condition). The result in this study indicated that the central-tendency bias is happened only at a later stage of processing the stimulus. The difference in central-tendency bias between in-view and the memory conditions supports that stimulus reconstruction uses the category-level knowledge, in the long run, this produces the bias in response. In keeping with previous study about the effect of category adjustment (Huttenlocher et al., 2000), this result suggests that central-tendency bias arises at a reconstruction stage of stimuli estimation to maximize the accuracy even though it causes bias in estimation.

Many of the studies on the category adjustment model rely upon experimental studies conducted in a laboratory, however, a number of studies now examine the category adjustment model using ecologically valid contexts. There have been questions as to whether the category adjustment model applies in more naturalistic contexts outside of laboratory experiments (Spencer & Hund, 2002; Engebretson & Huttenlocher, 1996; Huttenlocher et al., 1991). Typically, participants are asked to put a dot in the blank frame after exposing a dot inside a geometric frame in a location memory tasks (Huttenlocher et al., 1991; Sandberg, Huttenlocher, & Newcombe, 1996). The results of these studies show that biases toward the centers of spatial categories formed when individuals impose category boundaries within a geometric figure (Engebretson & Huttenlocher, 1996; Huttenlocher et al., 1991) and people are likely to be biased away for the midline and away from the edges of the perceptual frame (Huttenlocher et al., 1994). In addition, the greater uncertainty of fine-grained information, the larger biases (Huttenlocher et al., 1991; Huttenlocher et al., 1994). That is, in terms of spatial location, the category adjustment model suggests that individuals remembering a given location combine stimuli from different levels of the hierarchical structure of space. Therefore, the final estimate of location is hypothesized to be an optimal blending of a fine-grained metric estimate and coarser categorical information, represented by a category prototype (Holden, Curby, Newcombe, & Shipley, 2010).

As discussed previously, many researchers investigated biases toward the centers of spatial regions in the location memory tasks (e.g., Huttenlocher et al., 1994). However, it is unclear whether category adjustment model apply to memory of the natural locations, which contain irregularities in categories. Complex naturalistic environments are filled with semantic content as well as perceptual information (Holden, Curby, Newcombe, & Shipley, 2010). Research by Holden et al. support the validity of the category adjustment model in complex natural locations as well. The results clearly show that memory for locations within visually rich images is biased and that this bias tends to be in the direction of the category center, as predicted by the category adjustment model.

Although it is difficult to prove experimentally, it has been hypothesized that the cognitive processes underlying the category adjustment model may be part of the cognitive architecture and evolved, over time, as an adaptive strategy for increasing accuracy in estimation. Suggestive evidence for this possibility can be found in a study by Duffy, Huttenlocher and Crawford (2006), who examined whether the category adjustment model could also increase the accuracy in estimating among children. In this study, the authors investigated how categories affect stimulus estimation among 5- and 7-year-old children by exploring memory distortions that arose when they reconstructed tasks exhibiting in either peaked or uniform distribution condition (Duffy et al., 2006). This study was addressed three questions: (1) whether children use categories in estimation, (2) whether the shape or the bias curve vary by stimulus distribution and age, (3) whether response variability vary by distribution and age. Consistent with the predictions of the category adjustment model (Huttenlocher et al., 2000), both 5- and 7-year-old children constructed and used inductive categories when estimating stimuli (Duffy et al., 2006). Also, both age groups of children presented greater level of uncertainty in the peaked distribution than uniform distribution condition, because the stimuli near the region of average value were shown more frequently than the stimuli in the region of the extreme value. In addition, when it comes to the age in category adjustment model, 5-year-olds exhibited more bias and variability in their responses than 7-year-olds because younger children's memory is less precise than older children.

Taken together, the category adjustment model provides better understanding of the role categories perform in increasing the accuracy of memory and decrease the variability of estimates, providing a rational and adaptive process for maximizing the precision of memory under conditions of uncertainty.

Methodological issues involved in the category adjustment model

In order to explore category effects in memory, researchers have developed an experimental methodology known as the sequential stimulus reproduction task. Because everyone's experience with natural objects differs (i.e., you have a different set of exemplars of the category dog than, say, I do) it is useful to use artificial, transient categories that are produced in the laboratory. Much of the literature uses lines that vary in length, but others have used squares that vary in size or fish that vary in fatness. In these experiments, participants view and reproduce a series of stimuli randomly selected from a distribution of lines. On each trial, participants see a target line then after a delay see an adjustable reproduction lines that they shrink and expand to be the same length as their memory of the target line. In quite a short period of time, participants begin biasing responses toward the central value of the distribution they have seen (the formation of categories occurs rapidly). Over the course of the experiment, participants reproduce the entire distribution of lines. Although the number of lines used varies from experiment to experiment, most studies use between 100 and 200 lines, as having fewer leads to vague effects and more becomes too taxing for participants (Duffy, 2013, personal communication).

After collecting data from each trial on many subjects, the data is processed by subtracting the subject's response line, which is the measure of bias in responses. When examining experimental results of studies that use the category adjustment model, researchers generally look at the bias curve of responses. Figure 4 depicts a hypothetical bias curve. On the X-axis is stimulus values, going from, for instance, small to large or light to dark. On the Y-axis is bias, the extent to which people over or underestimate the stimulus value. Across many experiments, researchers find overestimation of small stimuli and underestimation of large stimuli.

Different studies have also varied in the way they present that adjustable line. Some studies always have the reproduction line have a starting length that is short, and other studies have the starting length that is long. It is unclear whether the starting length of the reproduction line influences responses. One hypothesis is that when the starting length is small, people's responses are underestimated, while when long, they are overestimated. This may be due to the starting length of the reproduction line interfering with the memory of the stimulus. If this hypothesis is true, then there should be a difference in intercept (but not slope) when the starting length is small as opposed to long, as depicted in Figure 5. This is the basis of Experiment 1a.

Another difference between studies concerns the velocity that the response stimulus expands or contracts. Some studies use reproduction stimuli that slowly expand and contract, others use stimuli that rapidly do so, and still others simply do not contain this information (including the original Huttenlocher, et al. (2000) study). It is unclear whether this factor could have an effect on resulting responses, although the category adjustment model would make a prediction about the speed of the response line. Specifically, when the response line moves slowly, it takes more time for participants to reach the desired length. With more time, the fine grain memory may degrade, and so people may rely more upon the category when the line moves slowly. If this hypothesis is true, then the slope (and intercept) of the bias curve should be steeper when the line moves slowly as opposed to quickly, as depicted in Figure 6.

The second experiment is related more directly to this issue of timing. Prior studies vary in the amount of time between the presentation of the target and the response line. Some studies use interstimulus response times of 1 second, others as long as 5 seconds. The effects of these differences are currently unknown, although the category adjustment model would predict that the longer the delay between target presentation and response, the greater the bias, for the same reason discussed previously.

Experiment 1

Participants: 43 undergraduates at Rutgers University in Camden participated to fulfill a course requirement.

Procedure and Design: In this and the following experiments, stimuli were horizontal lines that varied in length from 80 – 368 pixels in length and were 5 pixels wide, presented on a Windows desktop computer running E-Prime software. The distributions were 190 lines of 10 lines of each of 19 stimulus values starting at 80 pixels up to 368 pixels in 16 pixel increments. The lines were presented for 1.5 seconds, with a delay of 1.5 seconds, when the response line appeared. In the current experiment, there were four conditions. In the slow velocity/ small starting length condition, the response line started at 60 pixels and expanded and contracted at a rate of 16 pixels per second. In the slow velocity/ long starting length condition, the response line started at 388 pixels and expanded and contracted at a rate of 16 pixels per second. In the fast velocity/ long starting length condition, the response line started at 388 pixels and expanded and contracted at a rate of 32 pixels per second. In the fast velocity/ short starting length condition, the response line started at 60 pixels and expanded and contracted at a rate of 32 pixels per second.

After estimating the 190 lines, participants were debriefed and thanked.

Results

I first performed an analysis on the data combined to determine if there was an interaction between starting length and velocity. I first analyzed the data at the individual level by performing separate ordinary least squares (OLS) regressions with bias as the dependent variable and actual stimulus size as the independent variable. For every participant, actual size significantly predicted bias, suggesting that every participant had a negatively sloped bias curve. The individual t statistic for actual size predicting bias ranged from -1.4 through -22.02 with an average value of -9.70, suggesting that every subject showed the pattern of bias predicted by the category adjustment model.

I then performed two omnibus analyses of variance on the intercept and slope. This yielded a non-significant effect for condition $F(3, 39) = .025$, $p = .32$ for the intercept and $F(3, 39) = 1.6$, $p = .199$ for the slope. However, since the numbers of participants in each cell was quite small (approximately 10) I decided to perform separate analyses on the data collapsed across the two conditions and consider starting length and velocity separately.

Starting length analysis

To test at the individual level whether starting length influenced responses, the slopes and intercepts obtained by the individual regressions mentioned before were submitted to *t* tests. Neither the estimates of the slopes or the intercepts of the regressions showed a significant difference t (41) = -0.296 and 0.006, respectively, suggesting that the starting length of the reproduction line did not affect performance.

However, the analysis outlined above is atypical of those done by studies on the category adjustment model. More commonly, data is averaged across participants within a condition and then compared using multiple regression with dummy variables. To conduct this analysis, multiple regression using dummy variables was used, with bias as the dependent variable and actual size, a dummy variable for condition that tests for differences in the intercept, a dummy variable that tests for differences in slope, and with the small condition arbitrarily chosen as the reference category (coded 0). This analysis yielded a significant ANOVA F (3, 37) = 487.93, $p < .001$, $R^2 = .99$. There was a significant effect for the coefficient for the intercept (β = 53.12, $t = 22.50$, p < .001) actual size (β = -.272, $t = -27.66$, $p < .001$), but not the dummy variable intercept (β = .06, $t = .018$, $p < .98$) and neither for the dummy variable for the slope (β = 0.13, $t = .964$, $p < .34$). The fact that the slope and the intercept dummy variables were not significant suggests that the slope of the bias curve (i.e. the extent to which people adjust responses toward the center of the distribution) does not vary as a function of the starting length of the response line. These results are presented in Figure 7.

Velocity results:

I analyzed the velocity data at the participant level in the same manner as the starting length analysis. T tests compared the slope and intercept of those in the fast and slow moving line condition. This analysis yielded a marginally significant difference for the intercept t (41) = -1.95, p = .06 and a significant effect for the slope t (41) = 2.21, p = .03. This suggests that velocity had an effect on both the slope and the intercept of the regression lines.

To explore this finding further, I conducted an analogous analysis using group-level data. Multiple regression using dummy variables was used to analyze the data, with bias as the dependent variable, and actual size and dummy variables for the intercept and slope included, with the fast condition as the reference category (i.e., coded zero). This analysis yielded a significant ANOVA F (3, 37) = 467.70, p < .001, R^2 = .988. There was a significant effect for the coefficient for the intercept (β = 41.88, t = 17.49, p < .001) actual size (β = -.218, t = -21.91, p < .001), and the dummy variable intercept (β = 20.19, t = 5.94, p < .001) and for the dummy variable for the slope (β = -.084, t = -5.97, p < .001). The fact that the slope and the intercept dummy variables re significant suggests that the slope of the bias curve (i.e. the extent to which people adjust responses toward the center of the distribution) varied as a function of the velocity of the response line. Specifically the fact that the coefficient for the slope dummy variable was negative and the intercept dummy variable positive given that the fast condition was the reference category suggests that in the slow condition, the slope is even more negative, resulting in an intercept that intersects the y axis at a higher point. These results are presented in Figure 8.

Furthermore, one of the reasons why there was a difference in slope for the slow moving condition could be the length of the delay caused by the slowness of the velocity. I did the T test analyzing the length of time took the response for the slow versus fast condition. The average response time was for 4.4 seconds in the slow condition and 2.3 seconds in the fast condition. I did analysis and suggested that there is highly significant difference in response times (t (8168) = 36.17, p < .000) and this is because of the length of delay that may cause this changing bias. This investigated this Experiment 2 by altering the delay.

Experiment 2

Participants: 65 undergraduates at Rutgers University in Camden participated to fulfill a course requirement.

Procedure and Design: In this and the following experiments, stimuli were horizontal lines that varied in length from 80 – 368 pixels in length and were 5 pixels wide, presented on a Windows desktop computer running E-Prime software. The distributions were 190 lines of 10 lines of each of 19 stimulus values starting at 80 pixels up to 368 pixels in 16 pixel increments. The lines were presented for 1.5 seconds. After target presentation, there was a delay that varied with experimental condition. In the short delay condition, the delay between target presentation and response line generation was 1.5 seconds. In the long delay condition, the delay was 8 seconds. The starting length of the response line varied between 60 and 380 pixels in starting length (although the results of Experiment 1 suggest that this is not an important factor in influencing the results).

After estimating the 190 lines, participants were debriefed and thanked.

Results

Bias was calculated as in Experiment 1.

I first analyzed the data at the individual level by performing separate OLS regressions with bias as the dependent variable and actual stimulus size as the independent variable. For every participant, actual size significantly predicted bias, suggesting that every participant had a negatively sloped bias curve. The individual *t* statistic for actual size predicting bias ranged from -6.25 through -20.20 with an average value of -9.27, suggesting that every subject showed the pattern of bias predicted by the category adjustment model. I then compared whether the slopes for the two conditions (short delay / long delay) differed significantly by comparing the slopes and the intercepts of the regressions between the two groups using t tests. This analysis yielded a significant difference in the slope t (63) = -3.78 as well as the intercept t (63) = 3.7 between the two groups.

These data were submitted to a multiple regression analysis averaged across the group with dummy variables with bias as the dependent variable and a dummy variable coding for condition for both the slope and intercept of the regression line, with the short delay condition as reference category. (Although this analysis is redundant with the individual level analysis described above, this is the more common approach in the literature using the category adjustment model). This is how previous studies have the regression yielded a significant ANOVA, F (3, 37) = 501.97, $p < .001$, $R^2 = .99$. The coefficient for the intercept was $\beta = 62.56$, $t = 19.88$, $p < .001$, the coefficient for the stimulus size was $\beta = -.29$, $t = -22.33$, $p < .001$, the coefficient for the dummy variable coding for condition for the intercept was β = 28.40, $t = 6.38$, $p < .001$, and the coefficient for the dummy variable coding for condition for the slope was β = -.123, $t = -6.39$, $p < .001$. The significance of these last two coefficients suggests that delay affects the degree to which people introduce bias into their responses, with greater bias with the longer delay given that the short delay was the reference category in the regression and the coefficient was negative for the slope coefficient and positive for the intercept, suggesting a steeper curve in the long delay condition. These results are presented in Figure 9.

General Discussion

The present study addresses several methodological issues concerning aspects of the experimental design of tasks that explore category effects in stimulus estimation. The experimental method of sequential stimulus estimation has been used by a number of researchers across several laboratories, yet there has been little exploration of aspects of the procedure that might affect performance. The present study examines three of these aspects that from the perspective of the category adjustment model might affect performance: the starting length of the reproduction stimulus, the velocity of reproduction stimulus, and the length of delay between target presentation and response.

For the starting length of reproduction line, the category adjustment model makes no prediction, and the present study found no effect of this aspect of the design. Hence, future studies that test the category adjustment do not need to control for this aspect of the design of their experiments. However, the velocity of the reproduction stimulus, as well as the delay between presentation and response, do have an effect on resulting estimates. Specifically, for the slowly expanding/contracting reproduction stimuli line results in greater bias in the estimation of stimuli than the condition of fast expanding/contracting response stimuli. It is unclear whether it is the slow velocity of the line, or the increased length of time causes the increased bias, so Experiment 2 directly tested timing, finding that indeed, the delay between target presentation and response leads to a similar increase in bias. Hence, experimenters should be careful to have consistent timing in such studies, and should take care to use similar programs and set ups that maintain consistency.

In addition, for the longer the delay between presentation and response, participants appear to bias responses toward the center of the category to a larger extent than when there are shorter delays. This follows a prediction of the category adjustment model in that when the fine grain memory for a stimulus is imprecise (i.e., greater delay in the representation due to a longer delay) people would rely more upon the category prototype in constructing estimates of the stimulus.

Overall, longer delay and slow moving condition of response line exhibited greater variability in participants' responses than did shorter delay and fast moving conditions, suggesting greater memory inexactness for each stimuli. Recall that the model predicts that bias increases as a function of decreasing exactness in the fine-grain memories. Therefore, the pattern of bias observed across the two conditions suggests that participants are more likely to adjust their responses to a greater extent toward the prototype in order to compensate for the inexactness in their fine-grain memories for stimuli in longer delay and slow moving condition; on the contrary, participants introduce less bias into their response in shorter delay and fast moving condition because of their memory for individual stimuli is more accurate.

There are several directions for future research. The present study concern estimates of simple perceptual stimuli that vary along the different conditions. Different types of stimuli (e.g., spatial or social category) exhibit unique psychophysical properties that may influence the category value subjectively experienced as the distributions' center. The present study used lines that varied in length; future work on the estimation may explore how psychophysical properties of various stimuli affect estimation processes, particularly for stimuli with different Steven's Law exponent (Stevens, 1957). Furthermore, future researchers may be to explore whether the observed effect varies as a function of the delay between category induction and reconstructive judgment.

These findings are important in that many researchers rely upon sequential estimation tasks like the ones investigated in the present study, yet across laboratories there is very little systematic control over aspects of the design. Indeed, many studies even omit information about the factors explored in this study. These results suggest that certain aspects of the experimental design (delay and velocity) should be better controlled in future studies, or at least described in the methodology in order to better replicate the experiments in future investigations. Given the attention that replication has garnered in recent literature (Pashler & Wagenmakers, 2012; Pashler & Harris, 2012; Makel, Plucker, & Hegarty, 2012;

Bakker, Dijk, & Wicherts, 2012; Ferguson & Moiritz, 2012; Francis, 2012; Galak & Meyvis, 2012) it may be useful to include such information in order to improve upon psychological science.

In conclusion, studies using the serial estimation task need to take care to keep consistent timing and stimulus velocity in order to paint a clearer picture of how categories truly influence the estimates of inexactly encoded stimuli.

References

Bakker, D., & Dijk, & Wicherts, J. M. (2012). The rulers of the game called psychological science. *Perspectives on Psychological Science, 7*(6), 543-554*.*

Crawford, L. E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or Reconstruction? *American Psychological Society, 11*(4), 280-284.

Crawford, L. E., Huttenlocher, J., & Hedges, L. V. (2006). Within-category feature correlations and Bayesian adjustment strategies. *Psychonomic Bulletin & Review, 13*(2), 245-250.

Duffy, S., Huttenlocher, J., & Crawford, L. E. (2006). Children use categories to maximize accuracy in estimation. *Developmental Science, 9*(6), 597-603.

Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review, 17*(2), 224-230.

Engebretson, P. H. & Huttenlocher, J. (1996). Bias in spatial location due to categorization: Comment on Tversky and Schiano. *Journal of Experimental Psychology: General, 125*, 96-108.

Ferguson, C. J., Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science, 7*(6), 555-561.

Francis, G. (2012). The psychology of replication and replication in psychology. *Perspectives on Psychological Science, 7*(6), 585-594.

Galak, J., & Meyvis, T. (2012). You could have just asked: reply to Francis. *Perspectives on Psychological Science, 7*(6), 595-596.

Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.

Holden, M. P., Curby, K. M., Newcombe, N. S., & Shipley, T. F. (2010). A category adjustment approach to memory for spatial location in natural scenes. *Journal of Experimental Psychology, 36,* 590-604.

Hollingworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology and Scientific Methods, 7*, 461-468.

Huttenlocher, J., Hedges, L. V. (1994). Combining graded categories: Membership and typicality. *Psychological Review, 101*, 157-165.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and Particulars: Prototype effects in estimating the dates of events. *Psychological Review, 98*, 352-376.

Huttenlocher, J., Hedges, L. V., Lourenco, S. F., Crawford, L. E., Corrigan, B. (2007). Estimating stimuli from contrasting categories: Truncation due to boundaries. *Journal of Experimental Psychology: General, 136*, 502-519.

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., & Prohaska, V. (1992). Memory for day of the week: A 5+2 day cycle. *Journal of Experimental Psychology: General, 121,* 313-325.

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

Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology, 27*, 115-147.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*(6), 537-542.

Parducci, A. (1965). Category judgment: A range frequency model. *Psychological Review, 72*(6), 407-418.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*(6), 531-536.

Pashler, H., & Wagemakers, E. (2012). Editor's introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528-530.

Petzold, P., & Haubensak, G. (2004). The influence of category membership of stimuli on sequential effects in magnitude judgment. *Perception & Psychophysics, 66*(4), 665-678.

Sailor, K. M., & Antoine, M. (2005). Is memory for stimulus magnitude Bayesian? *Memory & Cognition, 33*(5), 840-851.

Sandberg, E. H., Huttenlocher, J., & Newcombe, N. (1996). The development of hierarchical representation of two-dimensional space. *Child Development, 67,* 721-739.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*(3), 153-181.

Tversky, B., & Schiano, D. (1997). Distortions in visual memory: A reply to Engebretson and Huttenlocher's comments on Tversky and Schiano. *Journal of Experimental Psychology: General, 126,* 212-214.