Running head: ANALYSIS OF RESPONSE TIME FOR PROTOTYPICALITY

# Analysis of Response Time for Prototypicality and the Effect of Ambiguity

Maya L. Uhl

California State University, Northridge

# Abstract

The prototypicality effect has been well tested using standard means analysis. However, response time distributions are known to be skewed, violating key statistical assumptions of normality and independence. Frequency, prototypicality and ambiguity were manipulated in a category decision task presented to 30 participants. The prototypicality effect was replicated and analyzed using standard means analysis, kernel density estimates, and bootstrapping methods. Response times indicated a prototypicality effect; however, the effect was less certain with the kernel density and bootstrapping techniques than with the means analysis. A significant ambiguity effect highlighted several areas for further study. These results indicate that standard means analysis is insufficient to describe the prototypicality effect and corroborates the recent trend to develop more sensitive analyses.

Analysis of Response Time for Prototypicality and the Effect of Ambiguity

According to prototypicality theory, categories are based on an ideal member (real or imaginary) to which all other instances are compared. No instances need be excluded from a category, but the worst instances will exist on the outer periphery of the category, far from the central prototype (Rosch, 1973, 1975; Rosch and Mervis, 1975). Rosch used response time and paper and pencil measures to show that participants tend to agree about the protoypicality of instances within a category. For example, most people seem to agree that only a few hues of red are most representative of the color red, while others are less representative.

However, dimensions of natural categories are continuous rather than discrete and appear to be hierarchically organized (Rosch, 1973, 1975), as larger categories often contain smaller ones. For example, the category "color" can be broken into subsets of warm and cool colors, which can be further divided into single color categories (red, blue, yellow) and further into hues of a single color.

The investigation of prototypicality has been an effort that seeks to "explain the categories found in a culture and coded by the language of that culture at a particular point in time" (Rosch, 1988, p. 312). Rosch proposed two general principles involved in the formation of categories: (1) categories function to provide maximum information with the least cognitive effort, and (2) categories allow us to perceive new information as structured rather than as arbitrary or unpredictable. Therefore, the maximum information with the least cognitive effort is achieved if categories closely map the perceived world structure.

Prototypicality has been shown to exert effects on several dependent variables including response time, learning speed, order and probability of output, and priming effects (Rosch, 1988).

However, it does not have clear implications for a particular processing model nor for a theory of cognitive representation of categories. The pervasiveness of prototypes in real-world categories and of prototypicality as a variable indicates that prototypes must have some place in psychological theories of representation, processing, and learning even though they don't themselves constitute any particular model of processes, representation, or learning (for a review see Rosch, 1988).

Difficulties with Rosch's (1973, 1975) work involve the stability of category effects. Homophones, homographs, frequency effects, and category width manipulations have produced varying results (Balota and Chumbley, 1984, Rosch, 1988). Rosch acknowledges that category boundaries are dynamic and depend largely upon context effects. For example, an individual who is contextually primed either by events in their life or in the course of a study (e.g., healthcare) will be more likely to interpret a homograph (e.g., invalid) as an exemplar of the primed category (e.g., an ill person) rather than another interpretation (e.g., not valid).

Research has most often centered on isolating prototypicality effects using independent variables to apply constraints, such as priming techniques (Rosch 1973, 1975), and narrowing category boundaries through normative and correlational studies (Uyeda and Mandler, 1980; Battig and Montague, 1969). Prototypicality effects have not been investigated by providing fewer rather than more constraints. The elimination of category ambiguity is generally sought during stimuli selection, resulting in the use of exemplars clearly rated as central or peripheral to the category. When category ambiguity has been an artifact of studies, it has often been explained as cultural or individual differences (Rosch, 1973, 1976; Beck, McCauley, Hershey, and Segal, 1988; Lin, Schwanenflugel, and Wisenbaker, 1990). Thus, category ambiguity has not itself been manipulated to ascertain any role it may play, or information it may yield.

The current study seeks to examine the effects of category ambiguity on prototypicality by introducing wide superordinate categories (living vs. nonliving) to compare with narrow categories (those used by Rosch, 1975; and rated by Uyeda and Mandler, 1980). The wide categories introduce ambiguity in two ways. First, superordinate categories have been shown to severely reduce priming effects found with basic-level categories – those that are more closely related (Rosch, 1975). For example, though participants may know the relation of dog as a living thing, they are faster to recognize the relation of dog as an animal. Secondly, exemplars have historically been rated in the context of narrow categories (e.g., birds, animals, toys, clothing, etc.), but have not been rated as to their goodness of representation to superordinate categories (e.g., living and nonliving). Another factor effecting prototypicality is word frequency, which has been shown to have a strong impact on lexical access (Balota and Chumbley, 1984). Because of these findings, word frequency will be manipulated as well.

The measure of interest is the response time of correct responses. As predicted by Balota and Chumbley (1984), a prototypicality effect is expected with narrow categories while a frequency effect is not expected. Thus, correct narrow trials using central exemplars are expected to require less response time than correct trials using peripheral exemplars. Conversely, a frequency effect is expected with wide categories while a prototypicality effect is not. Therefore, correct trials of high frequency words are expected to produce lower response times than low frequency words in wide category trials. A significant width effect is expected. Exemplars presented with narrow categories should yield lower response times than exemplars presented with the ambiguous wide categories.

Means analyses have most often been used to analyze differences in response time between central and peripheral exemplars. This has resulted in the generally agreed interpretation that central exemplars produce faster response times than peripheral exemplars, with no further investigation of the quantitative qualities of this relationship (Rosch, 1973, 1975, 1988; Rosh and Mervis, 1975; Collins and Loftus, 1975; Collins and Quillian, 1969). Indeed, the most common parameters estimated in response time analyses are the mean and variance of the distribution. However, response time distributions are not normally distributed; they are positively skewed (Van Zandt, in press; Heathcote, Popiel, and Mewhort, 1991). Thus, the basic statistical assumptions of normality and independence are violated. Further, because of skew of the distribution, the mean does not represent the most typical or likely value; rather, the most typical value is pulled in the direction of the skew. A similar problem exists with the variance. Outliers, a problem in any statistical analysis, are expected in response time data, and may represent the process of interest. However, the increase in variance associated with the presence of outliers reduces the power of the statistical tests performed (Van Zandt).

Van Orden, Moreno, and Holden (2003) suggest that response time distributions have a tendency to stretch as complexity increases, resembling pink noise ("an inverse power law on log/log scales," p. 52) rather than a log-normal distribution, suggesting a greater degree of complexity than a simple mean shift. However, complex response time tasks, such as categorization, tend to reveal shallower inverse power laws (Van Orden, Moreno, and Holden), and white noise, introduced as an artifact of unequal measurement intervals, makes pink noise more difficult to detect.

Van Zandt (in press) agrees that determining how response times are distributed is an important step toward isolating the process responsible for generating those times. Further, she suggests that distribution densities can be ascertained using a Gaussian kernel estimator. If the stretching effect is observed, a bootsrapping technique can be employed to normalize the data by repeatedly re-sampling the data set with replacement (Van Zandt, in press; Efron and Tibshirani, 1991). Once this is completed, the data sets are compiled to form a grand distribution. Using a skew statistic (Holden, 2002) both levels of the manipulated variable are combined to calculate a grand mean and standard deviation to which both levels of the manipulated variable can be compared – thus, showing each level's likely placement with respect to the other.

The present study will use a standard means analysis as well as a Gaussian kernel estimator to plot distribution shapes, and finally Van Zandt's bootsrapping technique (in press, MATLAB script, p. 73) will be used to investigate the nature of relationships between the levels of the dependent variables used in this study.

# Method

*Design*

The current experiment utilizes a 2 x 2 x 2 within-subjects factorial design. The first independent variable is the prototypicality of a member to its category and will consist of two levels: central and peripheral. The second independent variable is word frequency and will consist of two levels: high and low. The third independent variable is category width consisting of two levels: wide and narrow. Response times for correctly answered target trials will constitute the dependent variable.

*Participants*

Thirty-two psychology students at California State University, Northridge participated in the experiment. Eighteen of the participants were introductory psychology students who participated in partial fulfillment of course requirements. Eight of the participants were upper division undergraduate psychology students whose participation was voluntary, and the remaining six participants were psychology graduate students whose participation was also voluntary. All participants were right-handed native English speakers.

*Apparatus*

Stimuli were presented one at a time in random order on an IBM clone computer monitor using SuperLab Pro software (Cedrus, 1999). Participants responded by pressing keys on the computer keyboard which were labeled "yes" (the M key) and "no" (the Z key). Responses and response times were collected by the computer.

*Materials*

The stimuli consisted of a total of 672 trials composed of target words and categories selected from Uyeda and Mandler (1980). A total of 14 categories were selected (as detailed in the following paragraph): seven representing living things, and seven representing nonliving things. Each of the 14 categories was represented by 24 exemplars divided evenly into two sets: 12 true (e.g., bird: jay) and 12 false (e.g., bird: kale) category-exemplar sets. Further, each of the true and false category exemplar sets were also evenly divided into two different stimulus sets. The first set paired exemplars with categories listed in Uyeda and Mandler (narrow categories). The second set paired exemplars with the terms "living" or "nonliving" (wide categories).

The stimulus sets were constructed in stages. First, exemplars from each narrow category were selected only if they produced four exemplars representing one high frequency/central exemplar, one high frequency/peripheral exemplar, one low frequency/central exemplar, and one low frequency/peripheral exemplar. Thus, exemplars were chosen in yoked sets of four. High frequency words were rated above 10, while low frequency words were rated at 10 or below (as rated in Kucera and Francis, 1967). Central words were those with ratings below 2.8, while peripheral words were rated at 2.8 or above (as rated in Uyeda and Mandler, 1980). Further, once these criteria were met, an equal number of these yoked sets were selected to represent each of the wide categories (living and nonliving). A total of 120 exemplars met these criteria: 60 living and 60 nonliving (see Table 1 and 2). These 120 exemplars made up the target set to be used in later analysis. All were presented as true category-exemplar sets.

To insure the equivalency of target exemplars across salient dimensions, t-tests were conducted using alpha .05 for all comparisons. No significant differences were found between living and nonliving exemplars on the dimensions of word length, frequency ratings, or prototypicality ratings ($p > .39$). No significant differences were found between high and low prototypicality on the dimensions of word length and frequency ratings ($p > .50$). A significant difference between prototypicality ratings was found which served as a check for the prototypicality manipulation, $t(118) = 12.03$, $p < .0001$(high prototypicality $M = 2.06$, $SD = .46$; high prototypicality $M = 3.74$, $SD = .62$ (based on ratings from Uyeda and Mandler, 1980, 7-point scale). Finally, no significant differences were found between high and low frequency ratings on the dimensions of word length or prototypicality ratings ($p > .29$). Again, a significant difference was found between frequency ratings, $t(118) = 8.29$, $p < .0001$, which served as a check of the frequency manipulation (high frequency $M = 43$, $SD = 34$, low frequency $M = 3.4$, $SD = 3$, based on ratings from Kucera and Francis, 1967).

Forty-eight true filler sets were constructed to balance the number of true category-

Group 1.

The second experiment (Group 2) presented all 336 exemplars in opposite pairings as those in Group 1. In this way, the experiment was constructed so that half of the participants were presented with half of the target exemplars presented in conjunction with wide categories and half with narrow categories. The remaining participants were presented with the reverse exemplar-category combinations.

*Procedure*

The experiment was conducted in a room with several computer terminals. Participants were tested in groups of two to five. Five computers were used in the experiment. Before participants entered the room, half of the computers were set to run the Group 1 experiment, and half were set to run the Group 2 experiment. Each computer was used an equal number of times with both groups, and both graduate and undergraduate participants were equally represented in both groups.

At the beginning of the experiment, participants were greeted by the author and seated at a table facing a wall with a computer monitor and a keyboard positioned in front of them. Participants were told that instructions were presented on the computer screen (see appendix for instruction text), and that they could move through the instructions by pressing any key to continue to the next screen. After instructions were administered, an eight-item practice set was presented in which feedback was supplied for incorrect responses. Once the practice trials were completed, participants were instructed to stop and wait for the researcher. When all participants had arrived at this point, the researcher reviewed the instructions and answered questions about the procedure. Participants were instructed verbally (as well as on the computer) to respond as quickly as possible without making errors, and were informed that each person had a different version which would cause some to finish before others (a ploy to reduce distractions and comparisons between participant performances). Participants were asked to remain seated, still and quiet at the conclusion of the experiment to allow others to finish with minimal distraction. Participants were then told to press any key to begin the experiment.

Each experiment consisted of 336 trials presented in random order. Each trial began with a mask lasting 250 msec, followed by a category presented in blue type for 900 msec (e.g., fruit). Another mask was presented for 250 msec, followed by an exemplar presented in black type (e.g., apple). The participant's task was to determine whether the exemplar (shown in black) was an example of the category (shown in blue). If it was, they were instructed to press the key labeled "yes" with their right index finger (the M key). Conversely, if they thought that apple was not an example of fruit, they were told to press the key labeled "no" with their left index finger (the Z key). The trial ended when the participant pressed any key. All words were shown in the center of the computer monitor in bold Arial font, size 40, against a white screen. No capital letters were used.

The computer signaled the end of the experiment by stopping on a screen that instructed the participant to sit still and quiet until others had finished. This screen remained for three minutes, blocking any other computer activity. When all participants had completed the 336 trials, they were told the purpose of the experiment, allowed to ask questions, thanked, and excused. The entire experimental session lasted approximately 30 minutes.

## Results

Responses and response times (in msec) were recorded by the computer. Only correct responses to the 120 target exemplars were included in analysis. Responses exceeding 3500 msec were eliminated from analysis, as were anticipatory responses measuring less than 200 msec. Data from two volunteer undergraduate participants were thrown out as both stopped participating in the experiment and simply pressed one key throughout most of the experiment. Review of their reactions times also suggested their key presses were anticipatory responses. Two additional participants were recruited in such as way as to preserve the balance between groups regarding the type of student and equal use of computers.

Through the careful balancing of every aspect of the experiment between the two groups (prototypicality/frequency combinations, category width, number of exemplars per category, true and false presentations per category, equal use of computers for each group, equal use of groups during each experimental time period, equal representation of graduates and undergraduates in both groups), individual differences should be equally distributed throughout the planned comparisons, allowing a within subjects analysis. To verify the equivalency of participants in both groups, a t-test was conducted which revealed no significant differences at the alpha .05 level.

*Means Analysis*

A 2 (frequency) x 2 (prototypicality) x 2 (category width) repeated measure analysis of variance was conducted using an alpha level of .05 for all analyses.

A main effect of prototypicality was obtained, $F(1, 29) = 15.92$, $MSE = 28203.97$, $p < .001$, with central exemplars producing faster response times ($M = 1123.65$, $SE = 43.365$) than peripheral exemplars ($M = 1210.15$, $SE = 48.14$). A main effect of frequency was also obtained, $F(1, 29) = 4.70$, $MSE = 24653.30$, $p < .05$. Participants responded more quickly to high frequency exemplars (*M* = 1144.93, *SE* = 46.90) than low frequency exemplars (*M* = 1188.87, *SE* = 44.36). However, a problem was found with the accuracy of frequency ratings and therefore results of this variable are questionable.

The main effect of category width was also significant, $F(1,29) = 158.92$, $MSE = 43512.58$, $p < .001$. Participants responded faster when presented with narrow category exemplars (*M* = 997.16, *SE* = 42.08), than when presented with wide category exemplars (*M* = 1336.65, *SE* = 50.54).

As predicted, a significant interaction of prototypicality by width was found, $F(1,29) = 8.23$, $MSE = 29367.80$, $p < .01$. Response times were lowest when central exemplars were presented with their narrow categories (*M* = 967.83, *SE* = 40.67), followed by narrow/peripheral exemplars (*M* = 1072.14, *SE* = 47.28). However, when exemplars were presented in the context of wide categories, the prototypicality effect disappeared (*M* = 1325.12, *SE* = 51.69 for central exemplars, and *M* = 1348.17, *SE* = 54.46 for peripheral exemplars). No other interactions were significant.

Dunn-Bonferroni comparisons revealed significant mean differences between high and low prototypicality $t(29) = 86.50$, *SE* = 21.68, high and low frequency $t(29) = 43.94$, *SE* = 20.27, and narrow and wide category width $t(29) = 339.49$, *SE* = 26.93.

Figure 1 depicts cell means in all eight conditions. While not significant, the predicted pattern emerged in panel 2 with frequency and prototypicality in wide category presentation. However, frequency ratings thought to have come from Kucera and Francis (1967) were checked for accuracy and did not entirely match. Though the differences in ratings appeared minimal, they were sufficient to cast doubt on the goodness of the frequency manipulation. Therefore, an analysis of frequency effects would be tenuous at best, so frequency was collapsed across all further analyses.

*Kernel Density and Bootstrapping Analysis*

Prototypicality, category width, and their interaction were further explored using MATLAB. First, Gaussian kernel density plots were constructed to investigate distribution shapes. As shown in Figure 2, central exemplars presented with narrow categories yielded the highest mode, while peripheral exemplars presented with narrow categories yielded more variability; however, they both generated the expected skewed distribution shape. In contrast, both central and peripheral exemplars presented with wide categories were nearly indistinguishable from each other. Hence, the most striking differences appear between narrow and wide category widths.

To further explore the prototypicality effect, response times to central and peripheral exemplars presented with narrow categories were bootstrapped separately with 500 repetitions (Van Zandt, in press, p. 73). A skew statistic was calculated to plot both in respect to their combined grand mean. The result is shown in Figure 3. While the distributions overlap, the overlap is not complete. Hence, the prototypicality effect does not appear to be as robust as the standard means analysis would imply.

The same bootstrapping process was used with response times to central and peripheral exemplars presented with wide categories. As shown in Figure 4, both distributions overlap completely with only a small visible difference in variability between the two. Since prototypicality was not rated with wide categories, this pattern was expected. Its serves as a baseline and offers some evidence that Figure 3 may be capturing a true effect, however weak.

Finally, the effect of category width was explored using the same bootstrapping method. Figure 5 shows only a slight overlap between narrow and wide categories (with data collapsed across both frequency and prototypicality). Thus, width has emerged as the largest effect using these analysis techniques.

## Discussion

It was hypothesized that a frequency effect would be found with wide categories (as predicted by Balota and Chumbley, 1984), a prototypicality effect would be found with narrow categories (consistent with Rosch, 1973, 1975; Rosh and Mervis, 1975), and a main effect of width would emerge since the wide category introduced ambiguity.

The anticipated frequency by width interaction was not obtained. Though not statistically significant, the expected pattern was obtained, as Figure 1 shows. However, because frequency ratings were inadequately controlled, the lack of a frequency effect may simply be an artifact of inaccurate frequency ratings.

Consistent with expectations, a prototypicality by width interaction was obtained with means analysis. The means analysis revealed a significant difference between response times to central and peripheral exemplars when presented with narrow categories, but not when presented with wide categories. However, the shape of the distribution using a kernel density plot (Figure 2) shows a clear stretching effect. This provided evidence that more than a simple mean shift was involved, and brought into question the robustness of the prototypicality effect as viewed through simple means analysis. The bootstrapping technique was used to normalize the distributions of the levels of the manipulated variables by randomly re-sampling the data sets and compiling them into a grand distribution to represent the true placement of the prototypicality levels with respect to each other (Figure 3). This allowed a visual inspection of the overlap or distance between the two levels of prototypicality. Space between the two distributions would denote a significant difference between the levels. However, considerable overlap was revealed indicating either a questionable difference, or no difference at all. Since there is no generally agreed upon test of statistical significance for this finding (Van Zandt, in press), this discussion can only highlight the need for further research of the dimensions involved in the prototypicality effect. The current study was able to replicate Rosch's (1973, 1975; Rosch and Mervis, 1975) results as obtained by means analysis. The introduction of Van Zandt's bootstrapping technique calls into question the nature of the effect and demonstrates that something more than a mean shift is at work.

Figure 4 depicts response times to the same exemplars with wide category presentation. The distributions overlap completely leading to the conclusion that no significant differences exist between response times to central and peripheral exemplars in the wide condition. This was expected since prototypicality was not rated according to wide categories. Therefore, there would be no reason to anticipate a difference between central and peripheral response times. The addition of a wide category was intended to introduce ambiguity. From these results, it appears that ambiguity quickly mitigated the prototypicality effect.

The expected main effect of width was clearly obtained in means analysis, and further supported by the kernel density plot (Figure 2) and bootstrapping analysis (Figure 5). However, since a prototypicality by width interaction was obtained, a discussion of the main effect of width is speculative at best. However, such a discussion will be pursued which contains implications for further study.

By collapsing both prototypicality and frequency ratings, category width emerged as a

their narrow category responses. Therefore, the width effect obtained in this study may represent very different cognitive processes rather than a straightforward mitigation of the prototypicality effect using category width.

Future studies should attempt to mitigate the types of effects participants reported with wide category presentation, and instead isolate clear superordinate categories with their own scaled central and peripheral exemplars. It would be interesting to see if a stretching effect would occur with a more controlled category width manipulation, and whether bootstrapping would produce more or less distinct distributions.

Van Zandt (in press) suggests that the type of response time distributions produced in many studies can be described as *mixture distributions* (p. 58). Mixture distributions contain data whose parameters change over time, perhaps from fatigue or learning. As a result, the response times collected at the onset of the experiment may not resemble those collected toward the end. The data obtained in the present study likely falls into this category. There is no prescribed way of handling mixture distributions, making analysis very difficult. Consideration should be given to this problem in the future.

Another consideration for future study involves trimming. Because the process of interest may result in extreme scores, outliers are potentially interesting in the context of response time. Therefore, the elimination of data should not be undertaken (Van Zandt, in press; Heathcote, Popiel, and Mewhort, 1991). Heathcote, Popiel, and Mewhort assert that extreme values are more likely to reflect the process of interest when that process yields a skewed distribution than when it yields a symmetric one. Hence, trimming to correct a skewed distribution may actually remove evidence of the effect under study. Van Zandt cautions that when trimming is employed, both the trimmed and un-trimmed results should be presented so that readers are aware of potential artifact results.

In conclusion, the present findings question the prototypicality effect and suggest that a standard means analysis is insufficient to describe it. These findings corroborate a recent trend in the literature (Efron and Tibshirani, 1991; Heathcote, Popiel, and Mewhort, 1991; Van Orden, Moreno, and Holden, 2002; Van Zandt, in press) which highlights the need for more sensitive analyses than the standard means analyses can provide.

References

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10(3)*, 340-357.

Beck, L., McCauley, C., Segal, M.,, & Hershey, L. (1988). Individual differences in prototypicality judgments about trait categories. *Journal of Personality and Social Psychology, 55(2),* 286-292.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82(6),* 407-428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 241-247.

Cedrus (1999). SuperLab Pro (Version 2.0) [Computer software]. San Pedro: CA: Cedrus Corporation.

Efron, B., & Tibshirani, R. (1991). Statistical data analysis in the computer age. *Science, 253*, 390-395.

Heathcote, A., Popeil, S. J., Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the stroop task. *Psychological Bulletin 109(2)*, 340-347.

Holden, J. G. (personal communication, December 3, 2002).

Kucera, H., & Francis, W. N. (1967). *Computation analysis of present-day American English.* Providence, R.I.: Brown University Press, per Holden, J. G. (personal communication, October 7, 2002).

Lin, P. J., Schwanenflugel, P. J., & Wisenbaker, J. M. (1990). Category typicality, cultural familiarity, and the development of category knowledge. *Developmental Psychology, 26(5),* 805-813.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.

Rosch, E. H. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General, 104(3)*, 192-233.

Rosch, E. H. (1988). Principles of categorization. In A. M. Collins, & E. E. Smith (Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 312-322). San Mateo, CA: Morgan Kaufmann Inc.

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods and Instrumentation, 12(6),* 587-595.

Van Orden, G. C., Moreno, M. A.,, & Holden, J. G. (2002). A proper metaphysics for cognitive performance. *Nonlinear Dynamics, Psychology, and Life Sciences, 7(1)*, 49-60.

Van Zandt, T. (in press). Analysis of response time distributions. In J.T. Wixted (Ed.) *Stevens' handbook of experimental psychology (3rd Ediction)*. http://maigret.psy.ohio-state.edu/~trish/

Appendix: Computer-presented Participant Instructions

Screen 1. Press any key to begin.

2. In this experiment, a topic will appear in blue, for example . . .

3. furniture (shown in blue)

4. Then a word will appear in black, for example . . .

5. stereo (shown in black)

6. If "stereo" is a type of "furniture," press "yes" . . .

7. If not, press "no" . . .

8. Let's try a few.

9. Place your right index finger on the "yes" key . . .

10. Place your left index finger on the "no" key.

11. You only need to respond to the black words.

12. Press any key to begin.

13. (Eight practice trials given with feedback when response was incorrect.)

14. Please STOP HERE and wait for the experimenter.

(Experimenter reviews procedure, answers questions, and instructs participants to press any key to begin the experiment.)

15. We're ready to start the experiment.

16. Place your right index finger on the "yes" key.

17. Place your left index finger on the "no" key.

18. Respond as quickly as you can without making errors.

19. Press any key to begin the experiment.

Table 1

*Target exemplars yoked across living categories.*

| **Category** | **High Frequency High Typicality** | **High Frequency Low Typicality** | **Low Frequency High Typicality** | **Low Frequency Low Typicality** |
|---|---|---|---|---|
| Living | | | | |
| Four-footed animal | Dog | Buffalo | Goat | Rat |
| | Cow | Rabbit | Donkey | Camel |
| | Horse | Elephant | Tiger | Mouse |
| Bird | Eagle | Chicken | Crow | Pheasant |
| Body Parts | Legs | Shoulders | Toe | Ankle |
| Clergy | Bishop | Rector | Clergyman | Friar |
| | Minister | Missionary | Archbishop | Deacon |
| | Pope | Monk | Nun | Parson |
| Fruit | Orange | Fig | Banana | Prunes |
| | Lemon | Lime | Peach | Raisin |
| | Apple | Berry | Apricot | Melon |
| Trees | Walnut | Ash | Cedar | Dogwood |
| | Oak | Palm | Elm | Willow |
| Vegetables | Pea | Peppers | Carrot | Turnip |
| | Corn | Onions | Spinach | Kale |

Table 2

*Target exemplars yoked across nonliving categories.*

| **Category** | **High Frequency High Typicality** | **High Frequency Low Typicality** | **Low Frequency High Typicality** | **Low Frequency Low Typicality** |
|---|---|---|---|---|
| <u>Nonliving</u> | | | | |
| Carpenter's Tools | Nails | Knife | Hammer | Pliers |
| Clothing | Jacket | Tie | Blouse | Scarf |
| | Shirt | Belt | Trousers | Girdle |
| | Dress | Hat | Pants | Vest |
| Earth Formations | Mountain | Desert | Glacier | Gorge |
| | Island | Lake | Cave | Ravine |
| | Valley | Ocean | Volcano | Cavern |
| Musical | Drum | Bass | Banjo | Harp |
| Instrument | Violin | Viola | Flute | Tuba |
| | Guitar | Horn | Trumpet | Bugle |
| Reading Material | Newspaper | Poem | Paperback | Pamphlet |
| Sports | Football | Fishing | Hockey | Archery |
| | Baseball | Golf | Soccer | Fencing |
| Weather Phenomenon | Snow | Cold | Typhoon | Rainbow |
| | Storm | Clouds | Tornado | Sleet |

Figure Captions

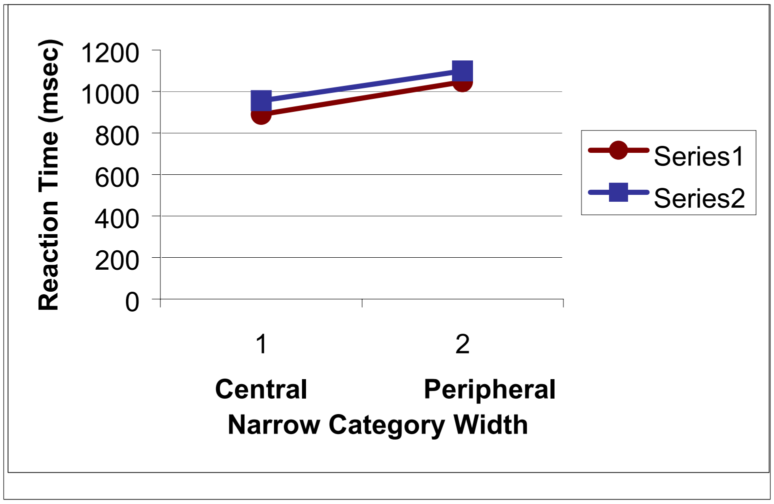
*Figure 1.* Cell means for all eight conditions as a function of category width, frequency and prototypicality. Panel 1 shows prototypicality with frequency in narrow category presentation, and panel 2 shows them with wide category presentation. Series 1, shown in red, denotes high frequency exemplars and Series 2, shown in blue, denotes low frequency exemplars.

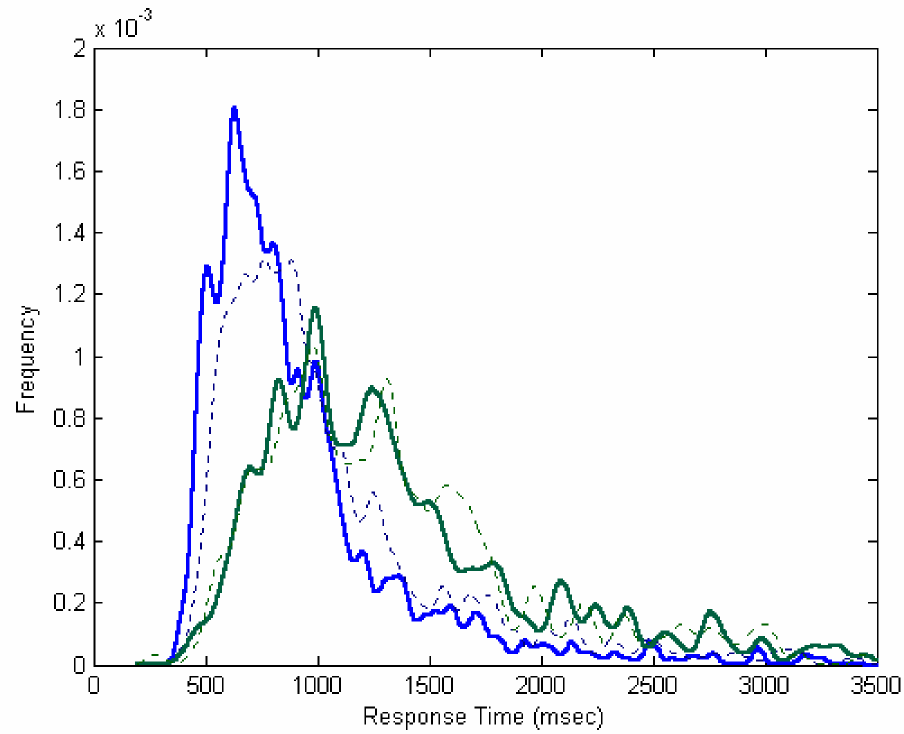
*Figure 2.* Distribution shapes of the prototypicality effect in narrow and wide category presentation. Central exemplars appear as solid lines, peripheral exemplars appear as dashed lines, narrow category presentation is denoted with blue, wide category presentation is denoted with green.

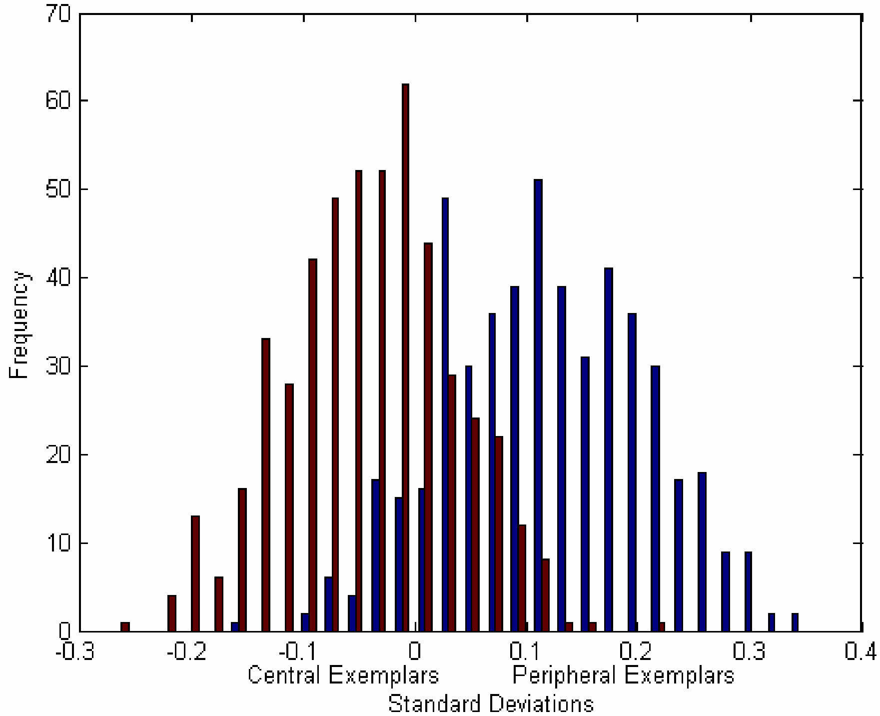
*Figure 3.* Bootstrapped distributions of the effect of prototypicality within narrow categories. Central exemplars appear in red and peripheral in black.

*Figure 4*. Bootstrapped distributions of the effect of prototypicality within wide categories. Central exemplars appear in red and peripheral in black.

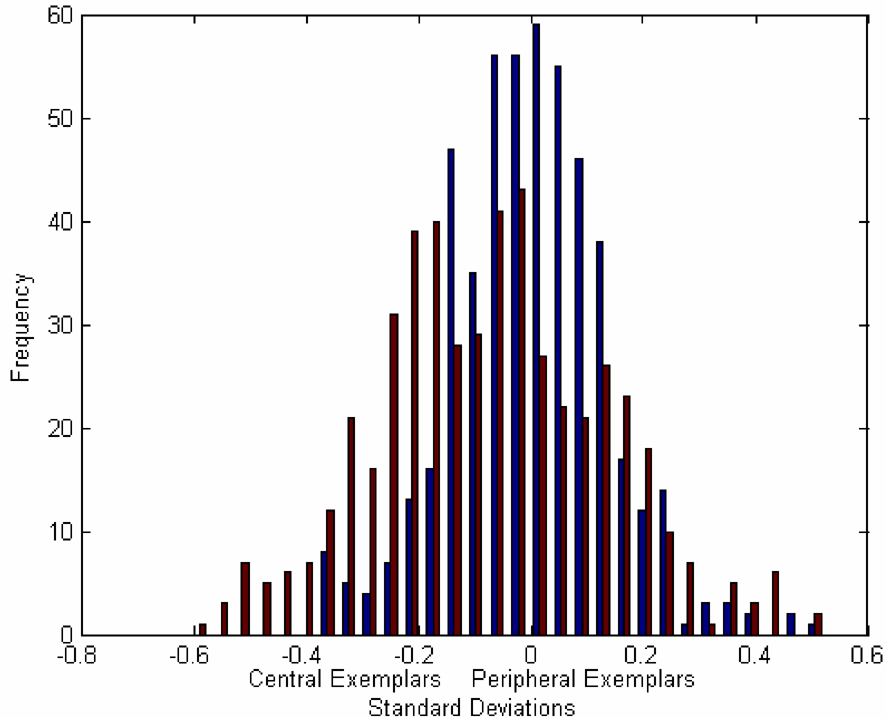
*Figure 5*. Bootstrapped distributions of the effect of category width (prototypicality and frequency collapsed). Wide exemplars appear in red and narrow appear in black.


Reaction Time (msec)
1200
1000
800
600
400
200
0
1
2
Central
Peripheral
Narrow Category Width
Series1
Series2

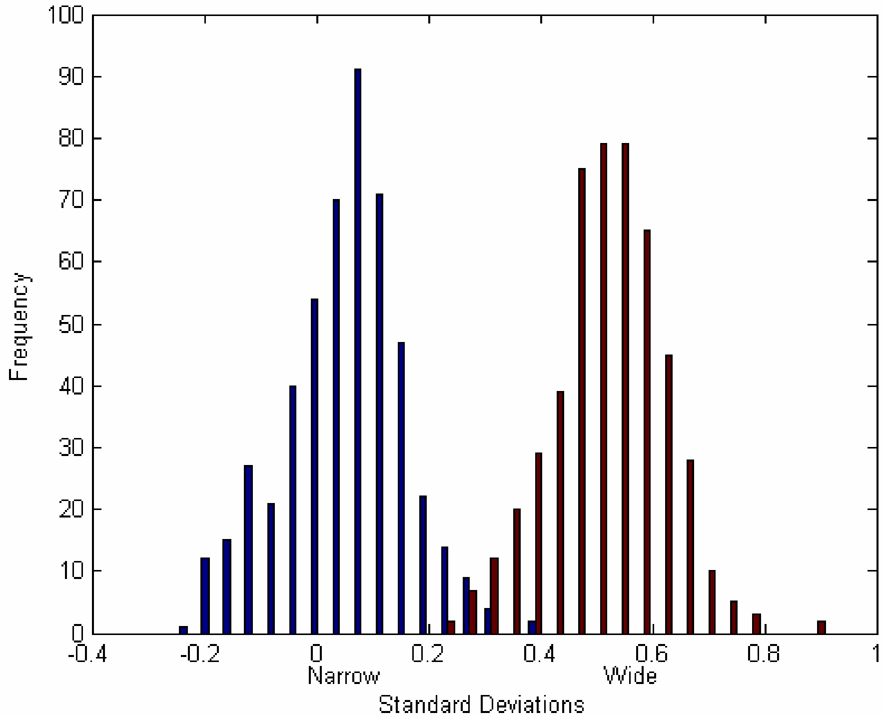

x 10-3
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Frequency
0
500
1000
1500
2000
2500
3000
3500
Response Time (msec)



70
60
50
40
30
20
10
0
Frequency
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
Central Exemplars
Peripheral Exemplars
Standard Deviations



60
50
40
30
20
10
0
Frequency
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
Central Exemplars
Peripheral Exemplars
Standard Deviations



Frequency
100
90
80
70
60
50
40
30
20
10
0
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Narrow
Wide
Standard Deviations