

INVITED ADDRESS

**DETERMINANTS OF SPEEDED CATEGORIZATION
IN NATURAL CONCEPTS**

Timothy VERBEEEMEN and Gert STORMS

University of Leuven

Tom VERGUTS

Ghent University

In this paper we describe two experiments in which the effects of typicality information from a target and a contrasting category on reaction times in a speeded categorization task were investigated. Experiment 1 elaborated on findings by Verbeemen et al. (2001) who found an effect of feature- and exemplar-based predictors from the target category, but virtually no effect of predictors from contrasting categories, in a wide range of categories. To avoid restrictions in the amount of relevant typicality information covered by these predictors we took a more direct approach. With directly rated typicality as a predictor for reaction times we obtained identical results. Experiment 2 expanded the set of concepts studied to superordinate animal concepts, again with typicalities as predictors. In these concepts we did find significant contributions of contrast categories. Possible explanations are discussed.

Rosch's (1973) and Rips, Shoben, and Smith's (1973) seminal research on natural language categories has shown that exemplars that are more typical of a category are categorized faster than exemplars that are judged less typical. This so called "typicality effect" has been interpreted as indicating that stimuli are categorized by computing similarity to a prototype (e.g., Hampton, 1979) or to stored exemplars of the category (e.g., Medin & Schaffer, 1978) rather than by verifying critical attributes that define the category (see Medin & Smith, 1981). Consequently, the categorization process is considered as a probabilistic rather than an all-or-none process. In this paper, we will concentrate on determinants of response times (RT) in speeded categorization.

In an elaborate study, Larochelle and Pineau (1994) replicated the finding

The first author is the recipient of the 2002 BPS Best Thesis Award and is currently working as research assistant of the Fund for Scientific Research - Flanders. This project was sponsored in part by grant OT/01/15 of the University of Leuven research council to Gert Storms.

Correspondence concerning this article should be addressed to Timothy Verbeemen, Department of Psychology, University of Leuven, Tiensestraat 102, 3000 Leuven. E-mail: Timothy.Verbeemen@psy.kuleuven.ac.be

that typicality within a category has a significant effect on RT of positive responses, but not on negative responses, in a speeded categorization task. They also found that familiarity significantly predicted RT of negative responses, but not of positive responses¹. Both the typicality and the familiarity measures used in their study were derived from ratings.

Rosch and Mervis (1975), however, showed that the graded structure within a category correlates positively with feature overlap within the target category, but negatively with feature overlap within contrasting categories, where a set of contrasting categories is defined as a set of conceptually related categories at the same level of abstraction (e.g., *birds*, *fish*, and *mammals*). This implies that, as a category member becomes more typical, it will have more properties in common with the target category and fewer properties in common with contrasting categories. A *robin*, for example, is expected to share a lot of traits with other *birds* but very little traits with, say, *fish*. Compared to a *robin*, a *penguin* would be expected to have much less in common with other *birds* and much more in common with the contrast category *fish*. Following up on this finding, Verbeemen et al. (2001) further investigated this contrast category effect in typicality ratings and in speeded categorization tasks. They found that there was no systematic independent contrast category effect on within-category structure of superordinate and basic level natural language concepts. More specifically, they found that, with the target and contrast category predictors combined in a linear regression, RTs from a speeded categorization task were significantly predicted from both feature-based and exemplar-based predictors of the target category, but that similar predictors related to the contrast category did not contribute significantly in the prediction.

One might argue that the absence of an independent contrast category effect in the study of Verbeemen et al. (2001) is attributable to the indirect typicality measures used. The predictor variables used in their study were based on applicability ratings (i.e., the number of participants that judged a feature as belonging to a presented item) of features from the target and the contrast categories (for the feature-based measures) and on rated similarity towards stored exemplars of the target and contrast categories (for the exemplar-based measures). In both cases, features and exemplars for a particular category were generated by a different group of participants in an earlier task. It is conceivable that such indirect measures do not capture essential typicality-related information in the contrast category: If the most frequently generated features or exemplars capture only the information that is most

¹ The RT task they describe presents statements of the form "a robin is a bird" and therefore allows for positive as well as negative answers.

important in determining the typicality of category members but not of contrast category members, as was the case in Verbeemen et al., then it would make sense to examine whether typicality itself is not a richer source of information than the original measures when predicting RT. Typicality can indeed be rated for both the target and contrast categories, and does not require a previously generated set of features or exemplars that serve as the only basis of comparison. In other words, it is possible that, when judging an item's typicality for a contrast category, participants use information that is immediately relevant to differentiate items with respect to that category. The information highlighted in the predictors used by Verbeemen et al. (2001) may differentiate well for target category items, but not for contrast category items. Therefore, the present paper aims at investigating whether directly rated typicality (for the target and contrast category) yields results that are different from the indirect measures used in Verbeemen et al. (2001) regarding the contrast category contribution in predicting RT. Two experiments will be described that directly investigate this question.

Experiment 1

Following up on the speeded categorization studies reported in Verbeemen et al. (2001), we intended to study contrast category effects with a different set of predictors. It could be argued that the indirect, feature- and exemplar-based predictors of typicality Verbeemen et al. used highlight information that is more important for the target category than for contrasting categories. For example, it is plausible that, when rating typicality for a contrasting category, people do not only use information associated with the best exemplars of that contrast category (on which the exemplar-based contrast category predictor in Verbeemen et al. was based). Experiment 1 examines the hypothesis that the predictor variables used in Verbeemen et al. may have highlighted only the information that influences the graded structure of the category itself, whereas a contrast category influence may originate mainly from another part of the information available within a concept. Therefore, the predictors we will use here are directly rated typicalities for the target and contrast category.

Method

Participants. Twenty students and research assistants from the University of Leuven participated in the typicality-rating task for the contrast category.

Material. The categories studied and the exemplar stimuli used were taken

from Verbeemen et al. (2001). The categories they used were *vegetables* (with *fruits* as a contrast category), *fruits* (with *vegetables* as a contrast category), *sports* (with contrast category *hobbies*), and *kitchen utensils* (with contrast category tools). Each category consisted of the 24 most frequently generated exemplars for that category (Storms et al., 2000).

RTs from the forced-choice task were also taken from Verbeemen et al. (2001). Ten participants completed their experiment. They sat in front of a computer screen and were told that they would see a number of items belonging to two different categories. Participants were asked to respond as quickly and accurately as possible, indicating their response by one of two buttons, either left or right. Each category corresponded to one of the two buttons (sides), counterbalanced over subjects. After a training session, they were presented the names of the two categories on the corresponding side of the screen, followed by all 48 items (2 x 24 category members). A plus sign indicated a new item would appear on screen. Participants performed this task for all three category pairs (three, since *fruit* and *vegetables* are each others contrast categories and are presented only once). The order of the three sets was randomized over participants, as well as the order in which the exemplars for each set were presented.

Finally, typicality ratings for the target category were taken from Experiment 2 of Verbeemen et al. (2001). They were obtained by rating the typicality on a seven-point rating scale, ranging from -3 to 3. Twenty participants completed their rating task.

Procedure. Participants in the typicality rating task for the contrast category each received four lists (corresponding to the four target categories), each list containing the 24 target category members, with the contrast category printed in bold on top. The order of presentation of the lists was counterbalanced over subjects. In order to get comparable results to the target category typicalities (taken from Experiment 2 in Verbeemen et al., 2001), participants were given a seven-point rating scale.

Results and Discussion

Reliabilities. The reliability of the ratings was calculated by correcting the split-halves using the Spearman-Brown formula. The estimated reliability for typicalities towards the contrast category were .769 for *fruits* (with *vegetables* as contrast category), .876 for *vegetables* (with *fruits* as contrast category), .647 for *sports* (with *hobbies* as contrast category) and .897 for *kitchen utensils* (with *tools* as contrast category). The reliabilities of the typicality ratings towards the target category, taken from Verbeemen et al. (2001), were all .90 or higher.

Table 1. Regression Results, Experiment 1

Variable	B	t-statistic	p-value (2-sided)
<i>Fruit and vegetables</i>			
$R^2 = .300$			
Typicality (target)	-.547	-2.987	.007
Typicality (contrast)	-.101	-.553	.586
<i>Vegetables and fruit</i>			
$R^2 = .318$			
Typicality (target)	-.566	-3.126	.005
Typicality (contrast)	-.041	-.229	.821
<i>Sports and hobbies</i>			
$R^2 = .407$			
Typicality (target)	-.647	-3.777	.001
Typicality (contrast)	-.056	-.327	.747
<i>Kitchen utensils and tools</i>			
$R^2 = .255$			
Typicality (target)	-.451	-2.394	.026
Typicality (contrast)	.219	1.164	.258

Regression analysis. RTs were predicted from rated typicality for the target and for the contrast category. Regression results are shown in Table 1. Percentages of explained variance were .300, .381, .255 and .407 for *fruits* (with *vegetables* as contrast category), *vegetables* (with *fruits* as contrast category), *sports* (with *hobbies* as contrast category) and *kitchen utensils* (with *tools* as contrast category), respectively. These values are comparable to the ones discussed in Verbeemen et al. (2001) and are not uncommon for RT studies with natural language stimuli. Regression weights corresponding to the target category predictor (typicality) were all significant and negative, according to predictions. Weights corresponding to typicality for the contrast category were never significant, and for all but one concept (*sports*) also in the wrong direction (i.e., negative). This result yields even less evidence for a contrast category effect than the results in Verbeemen et al. (see their Table 4), where at least for one concept (*kitchen utensils*), the exemplar predictor had a significant positive weight. It is also interesting to notice that typicality for the target category was never significantly correlated with typicality for the contrast category.

The results of Experiment 1 show that a contrast category effect is not general, not even when direct typicalities are used as predictors. The lack of significance for the feature-based and exemplar-based predictors used in Verbeemen et al. (2001) is therefore highly unlikely to be caused by inadequate predictor variables. Even when asking subjects to rate typicalities for contrasting concepts directly, the resulting predictor variables do not yield contrast category effects.

Experiment 2

In Experiment 2, we wanted to expand the set of concepts we studied. No significant and independent contrast category effect was found in the speeded categorization experiment in Verbeemen et al. (2001) and in Experiment 1, where manmade artifacts, activities, and natural kinds (food and animate beings) were studied. However, the animate beings studied in Verbeemen et al., that is, insects and reptiles, had rather low reliabilities. They may also not be the most representative examples of animal concepts. In Experiment 2, we study the concepts *fish*, *birds* and *mammals*: These categories arguably consist of animals that are more representative of the everyday environment of the participants. Furthermore, due to the influence of familiarity on RTs in categorization processes (e.g., Larochelle & Pineau, 1994), we also wanted to include rated familiarity of the stimuli as a predictor variable.

The experiment consists of three parts: First, a group of participants generates a number of animals for each of the three classes (fish, birds, mammals). Then, these animals, and a few extra, are rated for typicality and familiarity by a second group. Finally, a third group of participants performs a speeded categorization task and their RTs are predicted from the typicalities and familiarities delivered by the second group.

Method

Participants. Ten people participated in the *exemplar generation task*, fifteen in the *typicality rating task* for the target category and six for the contrast category, and nine in the *familiarity rating task*. Twenty-three people participated in the *speeded categorization task*. All participants were either research assistants of the University of Leuven or first year psychology students that participated for course credit.

Materials. All three possible pairs of the semantic categories *fish*, *birds*, and *mammals* were studied in this experiment.

Procedure. In the *exemplar generation task*, participants were asked to generate exemplars for the categories *fish*, *birds*, and *mammals*. Although some of the generated animals were not appropriate for the category (e.g., "squid" for the category *fish*), all responses from the generation task were included in the rating tasks. Furthermore, some extra exemplars for the three categories that belonged to under-represented subclasses (e.g., *water birds*) were added to the exemplar set. In total, 40, 43, and 32 animals for the categories *fish*, *birds*, and *mammals*, respectively, were selected for the rating task, yielding a total of 115 animals.

In the first *typicality rating task*, participants were required to judge the

typicality of each of the stimuli in the target category (i.e., the category for which they were generated). They used a five-point scale (ranging from "Absolutely not typical" to "Very typical"). The participants also had a choice option "I don't know this animal". The same persons participated in the *familiarity rating task*, where they were asked to judge the familiarity of each of the stimuli in the total stimulus set (consisting of exemplars from the three categories that were studied). Standard instructions were given for the familiarity rating task and the participants used a six-point scale (ranging from "Never heard of it before" to "Very familiar, I see it every day"). The order of rating (i.e., familiarity and typicality) was randomized over participants.

A different group of participants was asked to judge the typicality of each stimulus for each of the three categories (fish, birds and mammals). Hence, these participants made $3 \times 115 = 345$ ratings. Note that this resulted in three scores for each rater-animal combination: A typicality rating for the target category (i.e., the category to which the exemplar belongs), and a typicality rating for both of the other categories (which can be considered contrast categories; see Verbeemen et al., 2001).

Taken together, familiarity ratings were obtained for every stimulus from 9 raters. Typicality for the target category was obtained from $9 + 6 = 15$ raters, and typicality for the contrast categories was rated by 6 participants.

Participants in the *speeded categorization task* performed three tasks, in each of which a number of stimuli (i.e., animal names) were shown consecutively. In each task, the participant was asked to judge the stimulus as belonging to one of two categories as fast as possible (i.e., a two-alternative forced-choice design). On each trial, one category (e.g., *birds*) was presented on the left of the screen and the other (e.g., *fish*) was presented on the right side. A plus (+) sign was presented halfway between the two. After an exponentially distributed waiting time², an animal name was presented in the center of the computer screen, thus replacing the plus sign. The participant was asked to press a left green-colored button on the keyboard if the stimulus belonged to the left category, and a right yellow-colored button otherwise. The response side (left / right) to category mapping was constant for each participant but randomized over participants. The order of presentation of categories and the order of presentation of objects within a category were also randomized over participants. Accuracy, RT, and waiting time before presentation for each trial were recorded. Only accuracy and RT will be discussed here.

² The exponential distribution is the only distribution in which waiting times are not predictable by subjects (e.g., Green & Luce, 1971). However, because our participants had to read and understand the presented word first, and not just respond to any stimulus as fast as possible, this is not a necessary manipulation in the experiments we describe.

Ten example stimuli were shown at the beginning of every concept pair. For example, ten fish and birds were presented as practice trials to familiarize the participant with the category to response side (left / right) mapping. The sets of example and test stimuli were nonoverlapping.

All stimuli for which at least one participant from the familiarity rating task noted that (s)he did not know the animal was removed from the test stimulus set. In other words, we restricted our stimulus set to those items that were well-known to all participants, ensuring that the data are clearly interpretable³. The number of stimuli presented in the three category pairs were 85 (for *birds / mammals*), 82 (for *fish / mammals*), and 93 (for *fish / birds*).

Results and Discussion

Reliabilities. Table 2 shows reliabilities for the familiarity and typicality ratings (for the target and the contrast category) for all three concept pairs that were studied. The reliability estimates were derived using the split-half method (with Spearman-Brown correction). The reliabilities can be considered sufficiently high.

Table 2. Reliabilities for the Three Sets of Categories and the Three Predictor Variables, Experiment 2

	Fish/Birds (n = 50)	Fish/Mammals (n = 55)	Birds/Mammals (n = 68)
Typicality (target)	.829	.729	.721
Typicality (contrast)	.911	.750	.711
Familiarity	.801	.849	.695

Accuracies. Mean accuracies for the *fish / bird*, *fish / mammal*, and *bird / mammal* studies were .853, .829, and .858, respectively. These data show that the task was rather easy, which is not surprising given the fact that animals that were unknown to at least one participant in the familiarity rating task were excluded.

Regression analyses. As the main dependent variable, the mean RT on correct categorizations only was taken. The reason for this is that it makes little sense to incorporate a RT if the categorization on that trial was not correct. Otherwise, no (outlier) RTs were removed. These mean RTs were predicted

³ A few animals (i.e., dolphin, orca and (sperm) whale for fish/mammals, and bat for birds/mammals) were generated by participants for the inappropriate category and were switched to their appropriate category, mammals, to ensure that all objects are treated as members of their correct biological category.

in a linear regression using rated familiarity, typicality for the target category, and typicality for the contrast category as predictor variables.

Table 3 shows the results of the regression analyses for each of the three concept pairs. Percentages of explained variance were 67.5%, 35.2%, and 56.9%, for *fish / birds*, *fish / mammals*, and *birds / mammals*, respectively. These percentages are rather high as compared to the percentages reported in Verbeemen et al. (2001). Furthermore, it is remarkable that familiarity had no statistically significant effect in the prediction of the RT. It is also remarkable that both typicality in the target category and typicality in the contrast category had a strong and independent predictive value in the regression analysis.

Table 3. Regression Results, Experiment 2

Variable	B	t-statistic	p-value (2-sided)
<i>Fish and birds</i>			
$R^2 = .675$			
Familiarity	.021	0.246	.807
Typicality (target)	-.285	-2.908	.006
Typicality (contrast)	.640	6.550	<.001
<i>Fish and mammals</i>			
$R^2 = .352$			
Familiarity	.027	0.226	.822
Typicality (target)	-.426	-3.618	.001
Typicality (contrast)	.317	2.590	.012
<i>Birds and mammals</i>			
$R^2 = .569$			
Familiarity	-.077	-0.930	.356
Typicality (target)	-.642	-7.207	<.001
Typicality (contrast)	.200	2.264	.027

The independent effect of the contrast category is opposite to the findings of Verbeemen et al. (2001, Experiment 5) and Experiment 1 where no effect of similarity to the contrast category on forced-choice categorization RT was reported. How can this discrepancy be explained? One possible explanation is that outliers in the predictor variables were more pronounced in the present study than in Verbeemen et al. and Experiment 1. However, it is doubtful that these outliers can completely account for the difference in the results: Removing the four most extreme outliers (with respect to the typicality in the contrast category) decreased the percentages of explained variance for *fish / birds* and *birds / mammals* and increased the corresponding percentage for *fish / mammals* (33.0%, 54.2%, and 47.8%, respectively).

It is possible to gain further insight into why some of the RTs are large. Figure 1 plots predicted RTs (based on the regression model described above) versus empirically obtained RTs, for each of the three concept pairs.

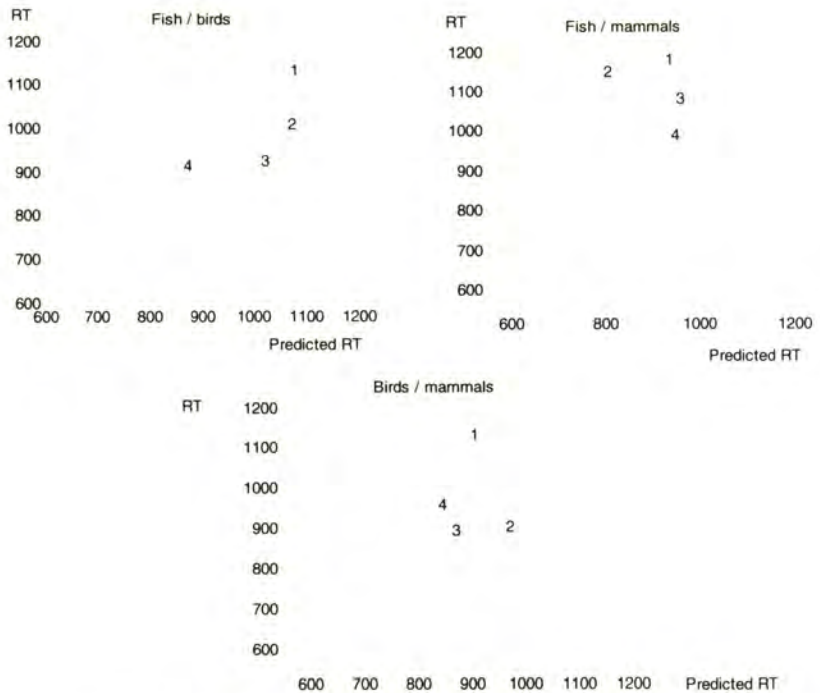


Figure 1: Predicted versus observed RT for all animal pairs, Experiment 2.

The best regression line is also shown.

In each of the three plots, four animals that both have high RTs and high predicted RTs (based on the typicalities) are marked. In the *fish / birds* plot, the stimuli labeled 1, 2, 3, and 4, are penguin, duck, swan, and goose, respectively. This makes sense in that all four are water dwelling birds. These birds are therefore presumably closer to typical fish than other, more typical birds are. In the *fish / mammals* plot, the stimuli labeled 1, 2, 3, and 4 are dolphin, orca, sperm whale, and whale, respectively. A very similar story can be told here: These are all mammals, but since they are water-dwelling as well, they are more similar to fish than typical mammals are, which increases RTs. Finally, in the *birds / mammals* plot, the labeled stimuli are penguin, bat, chicken, and duck, respectively. The high RT for bat is completely in line with the explanation given for the 4 extreme stimuli in the other two concept pairs. However, the fact that penguin and duck are outliers indicates that RT is not only high if an object has typical attributes of the contrast category. Instead, RT may be high whenever an object has atypical attributes whatsoever.

In conclusion, the data from Experiment 2 show that, at least for some categorization tasks, like the speeded categorization task described here, the similarity towards the contrast category can play a role in determining categorization responses. The results are in line with some findings reported in the literature but contradict others. First, we found a strong effect of typicality in the (target) category to which the object belongs on categorization response time. This finding is in line with many earlier studies both in semantic verification tasks (e.g., Casey, 1992; Larochelle & Pineau, 1994; Rips et al., 1973) and in forced-choice designs (e.g., Verbeemen et al., 2001, Experiment 5). However, the results of Experiment 2 also showed a strong (and independent) effect of typicality in the contrast category, which contradicts the results reported in Verbeemen et al. (2001, Experiment 5) and Experiment 1. We will elaborate on this in the General Discussion section. Finally, familiarity of the stimulus did not contribute significantly in the prediction of RT. The latter finding is in line with the results in the *positive* response condition in Larochelle & Pineau (1994).

General Discussion

In Experiment 1 of this paper, we reanalyzed the superordinate concept data from the RT study in Verbeemen et al. (2001). Direct ratings of typicalities, instead of indirect (exemplar- and feature-based) measures were used to predict the RTs from the forced-choice experiment. In this way, possible artifacts or inaccuracies in the indirect measures were avoided. The results showed a significant effect of the typicalities of the target category, but failed to show an effect for the contrast categories. Furthermore, typicalities for the target and the contrast category were never significantly correlated. In Experiment 2, we expanded the set of concepts in our study of contrast category effects to superordinate animal categories. Contrary to Verbeemen et al. and Experiment 1, we *did* obtain significant contrast category effects for three superordinate animal pairs. The results of Experiment 2 were thus opposite to those of Experiment 1.

The question then remains why we found a contrast category effect in Experiment 2, but not in Experiment 1 and in the studies reported in Verbeemen et al. (2001). One of the interpretations mentioned by Hampton (1998) states that, for a contrast category effect to be present, categories need to have a sufficient amount of exemplar overlap. This would imply that mainly artifact categories yield contrast category effects. The results of our studies described in the present paper and in Verbeemen et al. (2001), where both artifact and natural concepts were investigated (using typicality as well as RT as dependent variable), do not support this hypothesis. Another possible

explanation is that perhaps only animal categories show a contrast category effect. However, *insects*, with *reptiles* as contrast category, studied in Verbeemen et al. (2001), again with typicality as a dependant variable, yielded no significant effect⁴. However, a related hypothesis is that the categories need a sufficient amount of feature overlap. This is plausible as one considers some of the exemplars that contributed to the contrast category effect in Experiment 2. As discussed above in the analysis of regression plots (Figure 1), an animal such as *whale* has a number of features such as "is water dwelling" that are prominent in the other, contrasting category.

Yet another hypothesis, already suggested by Verbeemen et al. (2001), is that contrast category effects may be related to the degree of mastery of a category. As categories become better known, the contrast category effect may disappear. This is in line with the much higher accuracies for category decisions reported in Experiment 1 where virtually no errors (less than 2 %) were made, as compared to Experiment 2. In sum, taking this and Verbeemen et al.'s study together, reliable contrast category effects are found for some categories but not for others. However, exactly when such effects appear remains an issue for future research.

References

- Casey, P. J. (1992). A reexamination of the roles of typicality and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 18, 823-834.
- Edelman, S. (1999). *Representation and Recognition in Vision*. Cambridge, MA: The MIT Press.
- Green, D. M., & Luce, R. D. (1971). Detection of auditory signals presented at random times: III. *Perception and Psychophysics*, 9, 257-268.
- Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18, 441-461.
- Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, 65, 137-165.
- Larochelle, S., & Pineau, H. (1994). Determinants of response times in the semantic verification task. *Journal of Memory and Language*, 33, 796-823.
- Medin, D. L., & Shaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Medin, D. L., & Smith E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 241-253.
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and verification

⁴ It must be noted, however, that reliabilities of the predictor variables for these categories were somewhat lower.

- of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar based information in natural language categories. *Journal of Memory and Language*, 42, 51-73.
- Verbeemen, T., Vanoverberghe, V., Storms, G., & Ruts, W. (2001). The role of contrast categories in natural language concepts. *Journal of Memory and Language*, 44, 618-643.