

Running Head: SIMILARITY AND CATEGORY-SPECIFIC EFFECTS

Measuring the Influence of Similarity on Category-Specific Effects

Laura R. Shapiro

School of Life and Health Sciences, Aston University

Koen Lamberts

Department of Psychology, University of Warwick

Andrew C. Olson

School of Psychology, University of Birmingham

The research reported was carried out in partial fulfilment of the degree of PhD for the first author, who was supported by a studentship from the Biotechnology and Biological Sciences Research Council, UK.

Correspondence concerning this article should be addressed to L. R. Shapiro, School of Life and Health Sciences, University of Aston, Birmingham B4 7ET, UK.

E-mail L.R.Shapiro@aston.ac.uk Tel. 0121 204 4052

Abstract

There is evidence for both advantages and disadvantages in normal recognition of living over non-living things. This paradox has been attributed to high levels of perceptual similarity within living categories having a different effect on performance in different contexts. However, since living things are intrinsically more similar to each other, previous studies could not determine whether the various category effects were due to perceptual similarity, or to other characteristics of living things. We used novel animal and vehicle stimuli that were matched for similarity to measure the influence of perceptual similarity in different contexts. We found that displaying highly similar objects in blocked sets reduced their perceived similarity, eliminating the detrimental effect on naming performance. Experiment 1 demonstrated a disadvantage for highly similar objects in name-learning and name-verification using mixed groups of similar and dissimilar animals and vehicles. Experiment 2 demonstrated no disadvantage for the same highly similar objects when they were blocked, e.g. similar animals presented alone. Thus, perceptual similarity, rather than other characteristics particular to living things, is affected by context, and could create apparent category effects under certain testing conditions.

Measuring the Influence of Similarity on Category-Specific Effects

Studies investigating category-specific effects in normal processing have suggested that high levels of similarity within categories of living things can cause objects from these categories to be harder to process (e.g. Gaffan & Heywood, 1993; Humphreys, Riddoch & Quinlan, 1988; Lloyd-Jones & Humphreys, 1997; Shapiro & Olson, 2005; Vitkovitch & Tyrrell, 1995). However, recent studies have indicated that similarity has a different influence on performance, depending on the testing conditions (Gerlach, 2001; Lag, 2005; Lloyd-Jones & Luckhurst, 2002). For example, Gerlach (2001) found a *disadvantage* for living things in a difficult object decision task (i.e. requiring high levels of perceptual differentiation) when viewing conditions were optimal (so that perceptual detail could be used in the decision). In contrast, Gerlach found an *advantage* for living things when the object decision task was easy (i.e. requiring low levels of perceptual differentiation), and viewing conditions were sub-optimal (stimuli were presented in the periphery, reducing the amount of perceptual detail that could be used in the decision).

Gerlach (2001) suggested that the category-effect reversed under different testing conditions because the influence of perceptual similarity acts in opposite directions in different contexts. However, since he was unable to control perceptual similarity across the living and non-living items in his study, he could not be certain that it was the high levels of similarity within the living categories that caused the various category-effects he observed or other characteristics specific to living things. In order to determine whether the influence of perceptual similarity varies under different contexts, we employed a set of novel living and non-living stimuli from Shapiro and Olson (2005) that were matched

for perceptual similarity. These images were used to compare the effects of similarity when items were presented in mixed vs. blocked sets. The effects of similarity should be strongest when highly similar items are presented in a mixed set containing objects from both similar and dissimilar categories. For example, animals may appear highly similar to each other in the context of other dissimilar categories (such as furniture, vehicles, office stationery). In contrast, the items should be perceived as less similar when the other categories are taken away and we are able to focus on the features that differ, effectively increasing the separation between items in similarity space (Livingston, Andrews & Harnad, 1998). In sum, when highly similar items are presented alone (i.e. in a blocked set), these high levels of similarity should no longer have such a detrimental effect on naming performance.

Context and Similarity Effects

Gerlach's (2001) account of the reversal of category-effects under different testing conditions (see also Lag, 2005; Lloyd-Jones & Luckhurst, 2002) is consistent with the view that we adapt our perceptual processing according to task demands. There are two main literatures that relate to this account. Firstly, explicit judgments of similarity have been shown to change in different contexts, supporting the view of similarity as flexible and context dependent (e.g. Bassock & Medin, 1997; Goldstone, Medin & Halberstadt, 1997; Medin, Goldstone & Gentner, 1993; Tversky, 1977). Secondly, conceptual knowledge interacts with lower level perceptual processes, causing the discriminability of objects to vary according to task demands (e.g. Livingston et al., 1998; Gauthier, James, Curby & Tarr, 2003; Hayes, Taplin & Longstaff, 2002). We will

briefly outline these two literatures, before explaining how they relate to the current study.

Research on similarity judgements.

Parducci's (1965) range-frequency model of category judgement would predict that similarity judgements should be context-dependent. Parducci proposed that category judgements reflect a compromise between the range and frequency principles. This theory assumes that a participant judging stimuli on a fixed range (e.g. from small to large), would subdivide their psychological range into categories (i.e. small and large), and their judgements would be influenced by the stimulus frequencies. Ranks of the middle stimuli would increase when more small stimuli are added to the sample. In contrast, the ranks of the middle stimuli would decrease when more large stimuli were added to the sample. If this theory were applied to pair-wise similarity judgements within a fixed range from dissimilar to similar, ranks of the middle pairs would increase when more dissimilar pairs were added to the set, but ranks of the middle pairs would decrease when more similar pairs were added to the set. Thus, stimuli would be judged as less similar if they are presented in the context of many other highly similar items.

The extent to which similarity is context-dependent is central to the debate as to the power of similarity, as opposed to rules (or conceptual/semantic knowledge), to explain human categorisation. Sloman and Rips (1998) describe a continuum from "strong similarity", defined as an automatic perceptual process, through to "no similarity", a complete rejection of the usefulness of similarity as a construct. There are strong arguments against both the strong view and the rejection view, but there is convincing evidence for a weak view of similarity as a context-dependent process, where

the precise computation is based on a set of properties, that vary depending on the representations used as the input. Goldstone and Barsalou (1998) provide a powerful argument for this view of similarity, stressing the interplay between perceptual mechanisms and abstract thought. Support for this view is provided by evidence that judgements of similarity are mediated by: prior judgements (Lasaline, 1996; Wedell, 1996); exposure to metaphors (Kelly & Keil, 1987); separation of test items by a delay (Wedell, 1996) and the particular items presented in the comparison set (Bassock & Medin, 1997; Goldstone et al, 1997, Medin, Goldstone & Gentner, 1993, Tversky, 1977). For example, Goldstone et al. (1997) found that similarity judgements were influenced by the context within a single trial, so that a target could be judged to be more similar to stimulus A than stimulus B on one trial, more similar to B than C on another trial yet inconsistently be judged as more similar to C than A on another trial. In a second experiment, they demonstrated that the relative similarity of two alternatives to the target was influenced by a third alternative. A third experiment demonstrated that the dimensions that are used to make the similarity judgement are foregrounded or backgrounded, depending on the set of displayed objects. Together, these studies suggest that similarity judgements are not stable, and instead, vary according to the context in which the comparison is made.

Research on context and object discriminability.

The studies we have outlined so far have measured explicit similarity judgements, whereas we are specifically interested in whether context can influence the discriminability of stimuli used in a naming task. In terms of Parducci's (1965) range-frequency model, a group of similar items should be perceived as less similar to each

other when presented on their own (in a blocked stimulus set) than when they are presented together with other dissimilar items (in a mixed set). In a blocked set, the psychological distance between similar items would be expanded because the anchor point for what is considered dissimilar would change, and because the number of similar items would increase as a proportion of the whole stimulus set. The most dissimilar of the similar items would seem more dissimilar in a blocked set in the same way that a middle-sized stimulus seems larger when presented in a set composed of primarily small stimuli.

The most compelling evidence that context influences object discriminability comes from categorical perception: our perception that the distance between a pair of items that cross a category boundary is greater than between equidistant pairs of items within the same category (Liberman, Harris, Hoffman, & Griffith, 1957). Whereas initial investigations suggested that categorical perception was confined to innate perceptual categories of colour and sounds (e.g. Bornstein, Kessen, & Weiskopf, 1976; Eimas, Siqueland, Jusczyk, & Vigorito, 1971), recent research has demonstrated that categorical perception can be acquired (Ozgen & Davies, 2002, Goldstone, 1994; Livingston et al., 1998). Livingston et al. trained participants to categorise two-dimensional stimuli or pictures of real objects, and found that training induced within-category compression effects, indicating that learning changes the dimensional structure of perceptual similarity space. Similarly, Goldstone (1994) found that object discriminability could be mediated by changing the categorisation condition. Goldstone trained subjects to categorise objects, with one relevant dimension, and one irrelevant dimension. When subjects were then asked to make same/different judgements, subjects' discrimination sensitivity was greater when the pair varied on a dimension that was previously relevant in the

categorisation condition. These studies suggest that perceptual processes are influenced by our conceptual knowledge and can be manipulated according to task demands.

The cascade model (Humphreys et al., 1988) predicts that object discriminability critically affects naming performance. In this model, perceptual similarity influences processing during access to stored structural descriptions, when competing structural descriptions that are perceptually similar to the target are activated. More distracters would be activated for objects with many highly similar neighbours, and this would therefore critically affect the speed with which the correct description is chosen. Since processing is assumed to operate in cascade, these distracters continue to be activated during semantic processing, and so perceptual similarity directly affects semantic access and name retrieval. If the context in which objects are displayed causes fewer competing structural descriptions to be activated, then we would expect naming to be facilitated.

In sum, the research on similarity judgements suggests that object representations are flexible and change according to the context in which a comparison is made. In addition, research on the interaction between conceptual knowledge and perception indicates that context can impact on perceptual processing, and thus affect naming performance.

The Current Study

The current study investigated whether the influence of perceptual similarity on naming is mediated by the context in which stimuli are presented. We used the novel animal and vehicle images from Shapiro and Olson (2005) to measure the effects of similarity and category under different learning and testing conditions. For this set of images, perceptual similarity is matched across the living and non-living items, allowing

us to determine whether any change in perceived similarity among items is consistent for both living and non-living categories. Although our stimuli are novel, they are plausible examples of living and non-living things. Shapiro and Olson (2005) conducted extensive pre-tests to ensure that participants processed the images as examples of the intended categories. Fifteen participants almost always categorised the images correctly as animals and vehicles (as opposed to “either” or “neither”), with 95% accuracy for the similar vehicles, 97% accuracy for the dissimilar animals and 98% accuracy for similar animals and dissimilar vehicles. Ten other participants completed a nearest-neighbour questionnaire, where they were asked to state the first real object that came into their head on viewing each image for the first time. Only two responses (“teddy bear” for animal and “cake on wheels” for vehicle) out of 160 (ten participants responding to 16 images) were inconsistent with the intended category. Ten further participants completed a detailed semantic properties questionnaire immediately after they had completed training and name-verification for the images. Participants did not view the images when filling in the questionnaire, and were asked to respond from memory. Two naïve judges, who had not seen the images, scored 94% of the responses to vehicles as appropriate for an object from the vehicle category and 97% of the responses to animals as appropriate for an object from the animal category. The responses to vehicles that were coded as inappropriate can be explained by participants’ frequent use of the term “wings” for the similar vehicles (plane-type vehicles) which was coded as an animal response by the judges, and the responses to animals that were coded as inappropriate can be explained by participants’ frequent use of the term “shiny” to describe one of the animals (AD2-fish-type animal). Although these results indicate that our images are largely processed as

new members of existing animal and vehicle categories, it is unlikely that the entire object recognition process is the same as for examples of real objects. Thus, we will be cautious in our interpretation of any category effects we observe and will return to this issue in the discussion section.

One further issue is that time constraints may have a different impact on recognition for different types of object. Specifically, category effects may not be observable under unlimited time conditions, but may become apparent when there is pressure to respond quickly. To address this issue, we asked participants to respond under long and short deadlines to ensure that we did not miss categorical differences that are only apparent when responses must be made quickly.

Predictions.

Following Gerlach (2001), we hypothesise that high levels of perceptual similarity among living things creates apparent category effects in certain contexts. Since our living and non-living stimuli are controlled for perceptual similarity, we expect no main effects of category in either experiment. However, we predict that the effects of perceptual similarity will vary with context. Following Shapiro and Olson's (2005) observations about the influence of similarity on name-learning and name-verification, we anticipate that performance will be worse for highly similar objects when we present our stimuli in a mixed set, e.g. dissimilar animals alongside similar vehicles (Experiment 1). In contrast, when the images are presented in blocked sets, e.g. similar vehicles alone, the similar vehicles will appear less similar to each other when presented in isolation, resulting in a non-significant effect of similarity (Experiment 2).

Methods

Stimuli

The same set of 16 images was used for both Experiments 1 and 2, taken from Shapiro and Olson (2005) (see Appendix A). These images were associated with a set of 16 novel names, also taken from Shapiro and Olson (see Appendix B). The set of images consisted of 4 similar animals, 4 dissimilar animals, 4 similar vehicles and 4 dissimilar vehicles.

The animal and vehicle images were matched for similarity and familiarity, according to ratings collected by Shapiro and Olson. For the similarity pre-test, 15 participants rated all within-condition pairs of images (48 pairs) in random order using a scale from 1 to 5 (1 = a pair of images that looked nothing like each other, 5 = a pair of images that were almost identical). As expected, the similar animals and vehicles were rated as significantly more similar than the dissimilar animals and vehicles. There was no main effect of category, indicating that the animals and vehicles were equivalent in their within-category similarity, and there was no interaction between similarity and category. For the familiarity pre-test, 15 participants rated all 16 images one at a time using a scale from 1 to 5 (1 = a very unusual image, 5 = a very familiar image). There was no main effect of category indicating that the animals and vehicles were equally familiar, no main effect of similarity indicating that the similar and dissimilar images were equally familiar, and there was no interaction between category and similarity. In addition, pre-tests indicated that participants readily categorised the images as animals and vehicles and responded appropriately in a semantic properties questionnaire (see Shapiro & Olson for a full report of the pre-test results).

Participants

Forty-seven participants between the ages of 18 and 31 were tested, 15 participants undertook Experiment 1 and 32 participants undertook Experiment 2. All were undergraduate or postgraduate students at Birmingham University and either gained course credits or cash payments for participating.

Design and procedure

In Experiment 1, the images were presented in mixed groups (e.g. similar animals and dissimilar vehicles), and in Experiment 2, the images were presented in blocked groups (e.g. similar animals). Experiment 1 had three repeated measures factors: similarity (dissimilar/similar), category (animal/vehicle) and deadline (long/short). In each sitting, participants were presented with 8 images, either dissimilar animals and similar vehicles together or similar animals and dissimilar vehicles together. The experiment took place in 2 sessions each lasting approximately one hour, conducted at 2-day intervals. In the first session, participants learnt the names of eight images, followed by the name-verification tasks. In the second session, participants completed the training followed by name-verification tasks for the remaining eight images. Each set of eight images consisted of a living category and a non-living category, a similar category and a dissimilar category. The order of the groups in each session was counterbalanced between subjects. Experiment 2 had two between-subjects factors: similarity (similar/dissimilar) and category (animal/vehicle). As in Experiment 1, deadline was a repeated measures factor (long/short). Participants were presented with one group of 4 images and completed the experiment in one sitting. Participants were assigned to dissimilar animals, similar animals, dissimilar vehicles or similar vehicles.

Training.

An image was randomly selected from the set and displayed. Underneath the image was a selection of all possible names from the set in random order. The participant selected the appropriate name using the mouse. If the answer was incorrect, the correct name was displayed with the image. Accuracy was recorded during training. All images continued to be displayed in random order until the participant reached an accuracy criterion, at which point the training session terminated and participants were given a short break before continuing to the name-verification stage. The criterion in Experiment 1 was 8 consecutive correct responses for each image. The criterion in Experiment 2 was increased to 20 correct responses per image. These criteria were chosen to ensure that participants were equally familiar with the images in each experiment and achieved similar performance levels. These criteria resulted in similar overall accuracy in the training sessions for the two experiments (see Tables 1 and 2 in the Results section) and similar overall accuracy in name-verification (discussed on p 18-19 in the Results section).

Name-verification.

Name verification was used to measure naming speed and accuracy (as in Lawson & Jolicoeur, 1998). Name-verification allowed us to measure response latencies much more precisely than in a naming task using a voice key. This task also allowed the response deadline to be manipulated (as in Brockdorff & Lamberts, 2000 and Lamberts, 1995). After training, participants were given a name-verification task under time pressure. A name was presented for 800ms followed by an image, presented either for 600ms (long deadline) or 400ms (short deadline). Participants had to decide as quickly as

possible whether or not the name and the image matched. Participants were asked to respond before the image disappeared, using a button box to indicate whether the name and the image matched (i.e. correct name) or did not match (i.e. incorrect name). When a response missed the deadline, a buzzer sounded and the message “Too slow- Try to respond before the beep”, appeared. No other feedback was given. The long deadline task was given followed by the short deadline task, with a practice session before each task. Each test session consisted of every possible combination of non-matching name and image and an equal number of matching name-image pairs. In Experiment 1, there were 56 non-matching trials (8 images presented with each of 7 possible non-matching names) plus an equal number of matching trials (total 112 trials). In experiment 2, there were 12 possible non-matching trials (4 images presented with each of 3 possible non-matching names). However, to ensure that participants had time to become accustomed to the tasks in Experiment 2, we doubled the number of trials to make 24 non-matching plus 24 matching trials (total 48 trials). As shown in the Results section (see Figures 1 and 3), name-verification accuracy was roughly equivalent for the two experiments. This will be discussed further in the Results section of Experiment 2. RTs and errors were recorded.

Results

Experiment 1

Data from 14 participants were analysed. One participant's data were removed because of inappropriate responses in a semantic properties questionnaire conducted after this experiment (classification of all similar vehicles as living things, reported in Shapiro & Olson, 2005). The following analyses were conducted by participants (denoted by F1) and by image (denoted by F2). Analyses by participants allowed effects that are reliable

across participants to be observed, but would have been affected by a single image causing an inconsistent pattern of responses. In contrast, analyses by image allowed effects that are reliable across images to be observed, but would have been affected by individual participants responding in unusual ways.

Training.

Performance was significantly less accurate for the similar conditions: $F(1, 13) = 21.57$, $MSE = 25.67$, $p < .001$; $F(1, 12) = 14.52$, $MSE = 13.55$, $p < .001$. The means are shown in Table 1. There was no significant effect of category, both $F \leq 1.40$, $p \geq .26$, and no interaction on the analysis by participant: $F(1, 13) = 1.06$, $p = .32$. However, the interaction was almost significant by image: $F(1, 12) = 4.15$, $MSE = 13.55$, $p = .06$. As shown in Table 1, there was a trend for the difference between similar and dissimilar conditions to be less for the vehicles than the animals. Since the trend for a category by similarity interaction was observed only in the analysis by image, we can assume that this effect was reliable across images, but was not consistent across participants and therefore may indicate that a sub-set of participants were using a different learning strategy. In fact, two participants showed unusually poor performance for the dissimilar vehicles, relative to other conditions. One participant scored 75% for the dissimilar vehicles, compared to 86% for the similar animals (the other condition learned in the same training session). Another participant scored 81% for the dissimilar vehicles and 86% for the similar animals. No other participant performed less accurately for the dissimilar vehicles than the similar animals. It is likely that these two participants decided to prioritise learning the harder condition (similar animals), and therefore focussed less on the easier condition (dissimilar vehicles) during training. No equivalent pattern of responding was observed in

the name-verification data for these participants (and the interaction was non-significant, see below). Thus, this strategy did not have a noticeable impact on name-verification performance.

Name-verification.

One participant's data were removed because of a computer failure during one of the name-verification tasks, so name-verification data from 13 participants were analysed. All incorrect and missed responses were removed before analysing the data. In addition, non-matching trials where the name and image came from different conditions (e.g. name of dissimilar vehicle and image of similar animal) were removed since the effects of category and similarity could not be separated for these trials. The remaining data were combined across matching and non-matching trials. Since our use of response deadlines obliged participants to respond within a particular time frame, we expect any variation in RTs to be minimal and the strongest effects should be observed in accuracy. Nevertheless, we also analysed RT data to check that the pattern of results was consistent. Extreme RTs, defined as more than 3 times the interquartile range from the upper or lower quartiles, were excluded from each participant's data. Again, F1 denotes the analyses by participants and F2 denotes the analyses by image.

The short deadline had the intended effect of decreasing accuracy and shortening response times. Accuracy was poorer at the short deadline, $F(1, 12) = 25.04$, $MSE = 22.30$, $p < .001$; $F(1, 12) = 70.23$, $MSE = 25.84$, $p < .001$. Since the short deadline obliged participants to respond faster, RTs were significantly faster, $F(1, 12) = 81.08$, $MSE = 1020.42$, $p < .001$, $F(1, 12) = 1381.79$, $MSE = 17.91$, $p < .001$. This pattern is shown in Figures 1 and 2.

There was a significant effect of similarity on accuracy and RT. Accuracy was significantly poorer, $F(1, 12) = 27.66$, $MSE = 12.16$, $p < .001$, $F(1, 12) = 14.36$, $MSE = 76.09$, $p < .003$ and RTs were significantly slower for the similar conditions, $F(1, 12) = 28.07$, $MSE = 90.59$, $p < .001$, $F(1, 12) = 16.70$, $MSE = 52.37$, $p < .002$. This pattern is shown in Figures 1 and 2. In addition, there was no effect of category on accuracy or RT, all $F \leq .44$, $p \geq .52$. There was a significant interaction between deadline, category and similarity on RT by image, $F(1, 12) = 4.77$, $MSE = 17.91$, $p < .05$, but not by participants, $F(1, 12) = .68$, $p = .43$. Figure 2 indicates that there was a trend for the effect of similarity to be greatest for the vehicles at the long deadline. However, since this interaction was only observed in the analysis by image, we can assume that there were not enough participants showing this pattern of responses for the effect to be significant by participants. The data from individual participants revealed that 5 out of 13 participants showed a larger similarity effect for the vehicles. In fact, two of these participants responded particularly slowly for the similar vehicles at the long deadline (mean 434ms and 436ms, the slowest mean RTs for any participant, in any condition). Thus, a subset of participants showed a greater similarity effect for the vehicles. There were no other significant interactions, all $F \leq 2.45$, $p \geq .14$.

In sum, these results replicate the findings of Shapiro and Olson (2005). We found clear effects of similarity on accuracy and RT, alongside non-significant effects of category. This pattern, as in Shapiro and Olson, was consistent across both long and short deadline conditions.

Experiment 2

In Experiment 2, we aimed to determine whether these strong effects of similarity observed in the mixed design of Experiment 1, were diminished when stimuli were presented in blocked sets. We conducted a post-hoc power analysis using GPOWER (Faul & Erdfelder, 1992) to assess the power of Experiment 2 to detect an equivalent effect of similarity, as found in Experiment 1. The effect size of similarity in Experiment 1 was large, $\eta_p^2 = 0.70$ for both accuracy and RT analyses. With 32 participants, the power of Experiment 2 to detect an effect of this size was 0.97. Thus, we had a 97% probability of correctly rejecting a false null hypothesis in Experiment 2.

Training.

On average, the number of responses per image needed to reach criterion was slightly less for Experiment 2 than Experiment 1 (33 responses per image in Experiment 2 compared with 43 responses per image in Experiment 1). However, accuracy over the whole training session was similar for the two designs (see Tables 1 and 2). The percentage of correct responses in training was analysed. Although the means shown in Table 2 indicate that accuracy was slightly poorer for the similar conditions, this did not reach significance, $F(1, 28) = 3.20$, $p = .09$, $F(1, 12) = 1.4$, $p = .26$. Unlike Experiment 1, there was no significant advantage for the dissimilar images in training. As in Experiment 1, there was no significant effect of category, both $F \leq 1.06$, $p \geq .33$, and no significant interaction, both $F \leq 1.11$, $p \geq .30$.

Name-verification.

As in Experiment 1, non-matching trials that crossed conditions were removed, and the remaining data were combined across matching and non-matching trials. All

incorrect and missed responses were removed before analysing the data and extreme RTs were excluded. Although we expected that learning a smaller group of images would lead to better overall performance, there was no clear overall advantage in accuracy for Experiment 2 (mean accuracy per participant, across both tasks, was 76% in both Experiment 1 and Experiment 2). However, if we compare Figure 3 with Figure 1, it is clear that participants in Experiment 2 were slightly more accurate than participants in Experiment 1 for the long deadline condition, and slightly less accurate for the short deadline condition. It is possible that the shorter testing session of Experiment 2 (48 trials, as opposed to 112) was adequate for participants to become accustomed to the long deadline, but did not allow participants to become as accustomed to the short deadline. Nevertheless, participants in Experiment 2 were performing well above chance levels at the short deadline. Although the mean accuracy for the short deadline task looks poor in Figure 3, this is because the scores represent the number of correct, non-missed trials out of all missed and non-missed trials. If we remove all missed trials from our calculation, and instead measure the number of correct non-missed trials out of all non-missed trials, it is clear that participants were performing very accurately, when they managed to respond in time, scoring, on average, 89% for the long deadline task and 79% for the short deadline task.

The only significant main effect was of deadline. Accuracy was significantly poorer for the short deadline, $F(1, 28) = 78.48$, $MSE = 67.27$, $p < .001$; $F(1, 12) = 112.54$, $MSE = 23.45$, $p < .001$ and RTs were significantly faster, $F(1, 28) = 137.87$, $MSE = 718.53$, $p < .001$; $F(1, 12) = 629.63$, $MSE = 77.99$, $p < .001$. This pattern is shown in Figures 3 and 4. Unlike Experiment 1, Experiment 2 revealed no significant

effect of similarity on accuracy or on RT, all $F \leq 2.22$, $p \geq .15$. Consistent with Experiment 1, there were no main effects of category on accuracy or RT, all $F \leq 1.74$, $p \geq .21$. There was a significant interaction between deadline and similarity on RT by image, $F(1, 12) = 7.16$, $MSE = 77.99$, $p < .02$, but not by participants, $F(1, 12) = 1.23$, $p = .28$. As shown in Figure 4, performance was slightly faster, on average, for the similar conditions at the short deadline but the similar and dissimilar conditions were equal at the long deadline. In addition, Figure 3 indicates a trend for an interaction between similarity and deadline on accuracy. Specifically, the effect of similarity on accuracy appears to be greater for the long deadline task. However, the interaction between deadline and similarity on accuracy was non-significant, $F(1, 12) = 0.97$, $p < .33$; $F(1, 12) = 1.39$, $p < .26$. We can therefore assume that not enough participants showed this trend to give a significant interaction. The results from individual participants indicate that the mean score for the similar animals condition was dragged down by one participant who scored only 60% for the long deadline task, whereas the next lowest score for this condition in this task was 69%. Note that accuracy for this participant only dropped to 58% in the short deadline task, so their performance was fairly accurate overall.

There was also a significant interaction between category and similarity on RT by image, $F(1, 12) = 8.51$, $MSE = 233.32$, $p < .01$, but not by participants, $F(1, 28) = 2.46$, $p = .13$. As shown in Figure 4, there was a trend for RTs to be faster for the similar animals than the dissimilar animals, but slower for the similar vehicles than the dissimilar vehicles. However, this interaction was only significant by image, indicating that this effect was driven by a sub-set of participants. Comparing Figures 3 and 4, it is clear that this trend goes in the opposite direction for accuracy. Thus, there is a tendency for some

participants in the dissimilar-animals condition to prioritise accuracy over speed, but for some participants in the similar-animals condition to prioritise speed over accuracy, and vice versa for the vehicle conditions. However, the interaction on accuracy did not reach significance, $F(1, 28) = 1.23, p = .28$; $F(1, 12) = 3.21, p = .10$. All other interactions were non-significant, all $F \leq 3.19, p \geq .10$.

Discussion

Summary of results

The effects of similarity were consistent and strong in Experiment 1 (mixed design) but non-significant in Experiment 2 (blocked design). Thus, the context of the testing set can mediate the influence of similarity on name-learning and name-verification performance. However, we observed no effects of category in either experiment. This is in accordance with Shapiro and Olson's (2005) findings, and indicates that perceptual similarity, as opposed to semantic category is likely to be the main cause of apparent processing differences between living and non-living categories. Nevertheless, our study used artificial examples of living and non-living objects, and it remains possible that an influence of semantic category would be observed if examples of real objects were used. This issue is discussed in the Category Effects section, below.

Similarity Effects

The results from Experiment 1 add to existing evidence that high levels of similarity have a detrimental effect on normal naming performance (Gaffan & Heywood, 1993; Humphreys et al., 1988; Lloyd-Jones & Humphreys, 1997; Shapiro & Olson, 2005; Vitkovitch & Tyrrell, 1995). According to Humphreys et al.'s cascade model of visual object recognition, high levels of perceptual similarity affect processing during access to

stored structural descriptions. Since processing happens in cascade, perceptually similar structural descriptions that are activated along with the target continue to be activated during semantic processing, and therefore impact on the ease with which the target is identified and named. The aim of the current study was to investigate whether the influence of similarity was mediated in different contexts. Specifically, we investigated whether blocked presentation caused objects to be perceived as less similar, causing fewer highly similar structural descriptions to be activated, allowing the target description to be accessed more quickly and facilitating the naming process.

Context Effects

Previous investigations of similarity and context have focussed on how judgements of similarity between objects are mediated by the context of the comparison episode (Bassock & Medin, 1997; Goldstone et al., 1997; Medin, et al., 1993; Tversky, 1977). Our aim was to investigate whether these context effects impact on object discriminability, mediating the effects of perceptual similarity on object naming. We anticipated that participants' perception of the similarity between objects could be manipulated through changing the context of the testing set. This hypothesis was based on evidence of an interaction between conceptual and perceptual processes (Livingston et al., 1998; Gauthier et al., 2003; Hayes et al., 2002) and in particular, evidence that object discriminability changes according to task demands (Goldstone, 1994). In addition, we predicted that these changes in the discriminability of items should impact on the naming process (according to the cascade model, Humphreys et al., 1988). We investigated context effects by comparing participants' naming performance for the same images, presented in mixed vs. blocked sets.

In Experiment 1, when participants were trained to learn names for a set of similar novel objects and another set of dissimilar novel objects together, their name-learning and name-verification performance was worse for the set of similar objects. These effects of similarity in the mixed design experiment were consistent and strong. However, in Experiment 2, the similar sets of objects no longer had a detrimental effect on naming when similar and dissimilar images were presented separately. Participants were apparently able to factor out common features and concentrate just on the dimensions that were important for discriminating object identity (effectively spreading exemplars over a larger psychological range, as predicted by Parducci, 1965). These findings provide support for the weak view of similarity as a context-dependent process (Goldstone & Barsalou, 1998, Goldstone et al, 1997, Medin et al., 1993). Nevertheless, our findings do not undermine the role of similarity in category-specific effects. Instead, we argue that similarity has a powerful influence on naming, but that perceptions of similarity are context-specific.

Research on base rates in category learning suggests that the frequency with which items are displayed in the training set can affect the way attention is allocated to features (Kruschke, 1996). Specifically, frequent categories are encoded by their typical features and rare categories by their distinctive features. Thus, items are encoded differently, depending on their frequency, and this affects the perceived similarity among items. Similarly, in our study, participants may have allocated attention to features differently when items were presented in mixed or blocked sets. In the blocked design, the items had many common features that could be ignored, and therefore attending to the remaining crucial features that differed would have decreased the perceived similarity

between items. In contrast, in the mixed design, the images that share common features formed only a subset of the images in the set, making it impossible to attend to only a few distinctive features. Thus, when participants could not simply attend to a few distinctive features in the mixed design, similar items were harder to name.

An interesting interpretation of the mixed design results is that the range (Parducci, 1965) was determined by the entire stimuli set, rather than by each category in the set. If participants were able to subdivide their psychological range into animals and vehicles or similar and dissimilar, this could have eliminated any effects of similarity, even in the mixed design. However, this was not the case. The addition of dissimilar items alongside similar items meant that participants were unable to “block” the similar category and just focus on the features that changed. In fact, the number of dissimilar items necessary to eliminate the effects of blocking was very small. A set containing just 4 dissimilar and 4 similar items was sufficient to cause difficulties in processing the similar items.

The context of the testing set may have had a greater influence on the results because of our training and name-verification paradigm and the use of novel images. The extensive training session ensured that the objects in the testing set were much more likely to be brought to mind than other objects from long-term memory. In addition, similarity within a set is likely to have a stronger impact in a name-verification task than in normal naming. In the name-verification task, the participant was shown a name followed by an image and the task was to decide whether they matched. Since the names and images were always taken from the testing set, the participant was constantly making within-set comparisons and this may have encouraged confusions within the set over

confusions with other objects in long-term memory. Finally, the set of novel images would have been perceived as a coherent group of unfamiliar objects, separate from other real-life objects the participants would have regularly encountered. This may have increased the likelihood of confusions within the set relative to confusions with other real-life objects.

Category effects

We found no effects of semantic category in either Experiment 1 or Experiment 2 supporting the hypothesis that perceptual similarity could cause apparent processing differences between objects from living and non-living categories (e.g. Gaffan & Heywood, 1993; Gerlach, 2001; Humphreys et al., 1988; Lloyd-Jones & Humphreys, 1997; Lloyd-Jones & Luckhurst, 2002; Shapiro & Olson, 2005). In addition, we found that the blocked design eliminated the similarity effect for both the animal and vehicle conditions, demonstrating that perceptual similarity, as opposed to other characteristics particular to living things, is affected by context.

Nevertheless, because we used artificial examples of living and non-living things in our study, we cannot preclude the possibility that some influence of category exists and this influence could vary in different contexts. Although Shapiro and Olson (2005) provided evidence that participants processed our novel images appropriately in off-line tasks, we cannot be entirely sure that participants were processing the stimuli in the same way as real things during our naming task. Thus, a more certain way to measure processing differences would be to use pictures of real things. Although it would not be possible to perfectly match the stimuli for perceptual similarity, future studies could investigate whether blocked presentation improves performance for similar items from

both living and non-living categories. The findings would depend on how susceptible identification judgements are to a stimulus set. If each picture of a real object were only presented once, other stimuli in the set would be unlikely to influence the object recognition process. Instead, structural descriptions of objects from the participant's long-term memory store would be more likely to interfere with the process. In contrast, if each picture were repeated many times, the participant would be very likely to build up a "set" of objects in memory, and structural descriptions of objects from this set would be more readily activated and therefore more likely to interfere with the recognition process. In addition, using a name-verification task would also increase the chance of confusions within the stimulus set, since participants are constantly asked to compare pairs of objects from the set when they are presented with an image and a name in each trial. If we find that blocking real images diminishes the influence of similarity, this paradigm could overcome the problem of controlling for similarity and semantic category independently. Even if the living and non-living items were not matched for similarity, the effects of similarity in the mixed design would disappear in the blocked design, revealing any remaining differences in performance that were due to semantic category alone.

References

- Bassock, M., Medin, D. L. (1997). Birds of a feather flock together: similarity judgments with semantically rich stimuli. *Journal of Memory and Language* 36, 311–336.
- Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976) Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance* 2, 115-129.
- Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old new recognition judgments. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 77-102.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Faul, F. & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS [Computer program]. Bonn, FRG: Bonn University, Department of Psychology.
- Gaffan, D., & Heywood, C. A. (1993). A spurious category-specific visual agnosia for living things in normal human and nonhuman primates. *Journal of Cognitive Neuroscience* 5, 118-128.
- Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, 20 (3-6), 507 - 523.
- Gerlach, C. (2001). Structural similarity causes different category-effects depending on task characteristics. *Neuropsychologia*, 39 (9), 895-900.

- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*, 178–200.
- Goldstone, R. L., & Barsalou, L. (1998). Reuniting perception and conception. *Cognition*, *65*, 231-262.
- Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition*, *25* (2), 237-255.
- Hayes, B. K., Taplin, J. E., & Longstaff, M. (2002). Prior knowledge and exemplar similarity in category learning: Further evidence for their integration. *European Journal of Cognitive Psychology*, *14* (4), 549 – 571.
- Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, *5*, 67-103.
- Kelly, M. H., & Keil, F. C. (1987). Metaphor comprehension and knowledge of semantic domains. *Journal of Metaphor and Symbolic Activity*, *2*, 33-51.
- Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 3-26.
- Lag, T. (2005). Category-specific effects in object identification: what is “normal”? *Cortex*, *41*, 833-841.
- Lassaline, M. E., 1996. Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 754–770.
- Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, *124*, 161-180.
- Lawson, R., & Jolicoeur, P. (1998). The effects of plane rotation on the recognition of brief masked pictures of familiar objects. *Memory & Cognition*, *26*(4), 791-803.

- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.
- Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 732 - 753.
- Lloyd-Jones, T. J. & Humphreys, G. W. (1997). Categorising pears and naming chairs: Category differences in object processing as a function of task and priming. *Memory & Cognition*, *25* (5), 606- 624.
- Lloyd-Jones, T. J., & Luckhurst, L. (2002). Effects of plane rotation, task, and complexity on recognition of familiar and chimeric objects. *Memory & Cognition*, *30* (4) 499- 510.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254-278.
- Ozgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, *131*, 477–493.
- Parducci, A. (1965). Category Judgment: A Range-Frequency Model. *Psychological Review*, *72* (6), 407–18.
- Shapiro, L. R. & Olson, A. C. (2005). Does normal processing provide evidence of specialised semantic subsystems? *Language and Cognitive Processes*, *20*(6), 697-724.

Sloman, S. A. & Rips, L. J. (1998). Similarity as an explanatory construct. *Cognition*, 65, 87- 101.

Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84 (4), 327- 352.

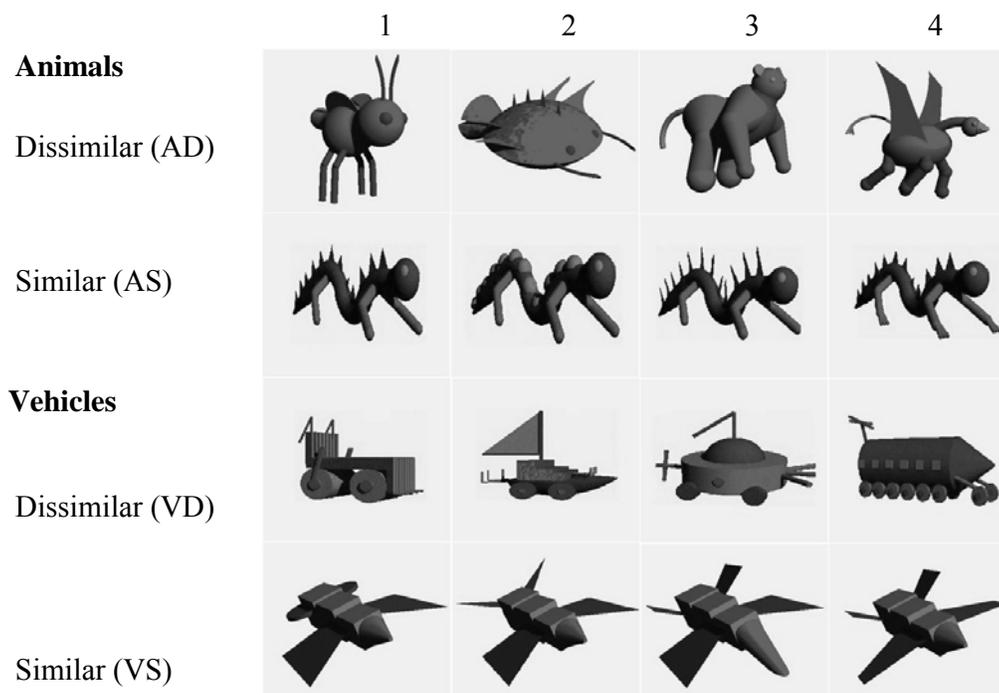
Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming.

Quarterly Journal of Experimental Psychology, 48A, 822- 848.

Wedell, D. H. (1996). A Constructive-associative model of the contextual dependence of unidimensional similarity. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 634- 661.

Appendix A

Dissimilar and Similar Novel Animals and Vehicles, Labelled AD1 to VS4 (from Shapiro & Olson, 2005)



Appendix B

Names for Images (from Shapiro & Olson, 2005)

Group 1	Group 2	Group 3	Group 4
Bruka	Shoni	Dronu	Flipi
Zidle	Kimba	Vulsa	Jutle
Denil	Pukid	Fulag	Gigip
Trolb	Glond	Chask	Skung

Table 1

*Means and SDs (Standard Deviations) for Accuracy in the Training Session of
Experiment 1*

	Mean % correct
Dissimilar Animals	88.72 (4.71)
Similar Animals	80.48 (9.30)
Dissimilar Vehicles	85.31 (7.37)
Similar Vehicles	80.98 (7.55)

Table 2

Means and SDs for Accuracy in the Training Session of Experiment 2

	Mean % correct
Dissimilar Animals	88.14 (9.35)
Similar Animals	80.68 (8.21)
Dissimilar Vehicles	87.62 (5.84)
Similar Vehicles	85.69 (5.60)

Figure Captions

Figure 1. Mean name-verification accuracy for each condition in Experiment 1, with standard error (S.E.) bars.

Figure 2. Mean name-verification RT for each condition in Experiment 1, with S.E. bars.

Figure 3. Mean name-verification accuracy for each condition in Experiment 2, with S.E. bars.

Figure 4. Mean name-verification RT for each condition in Experiment 3, with S.E. bars.

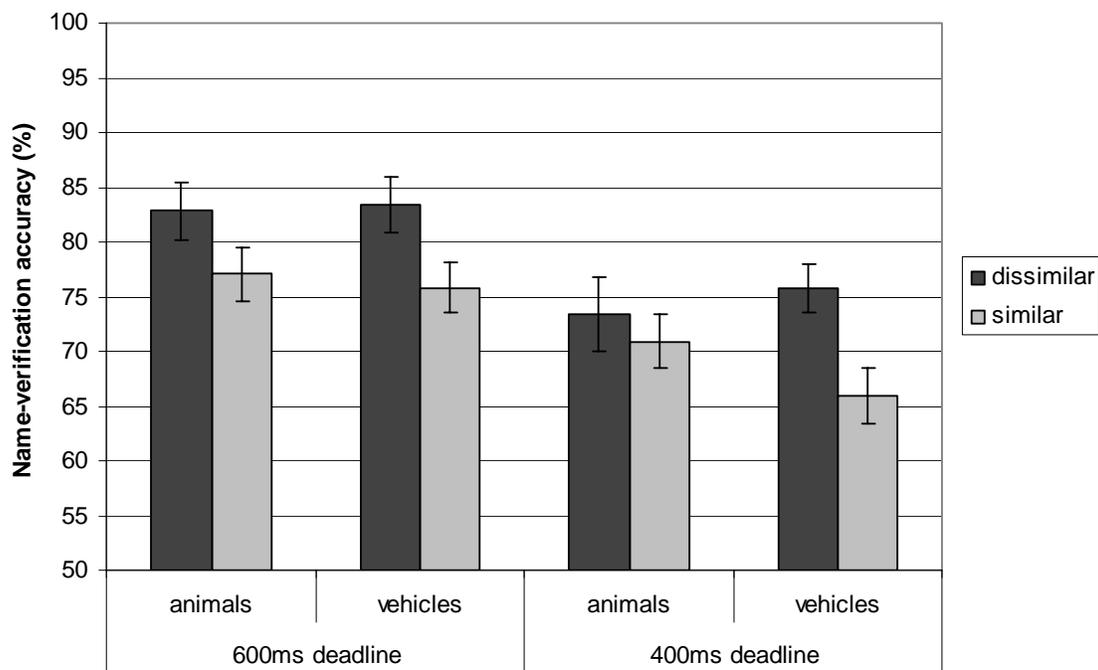


Figure 1

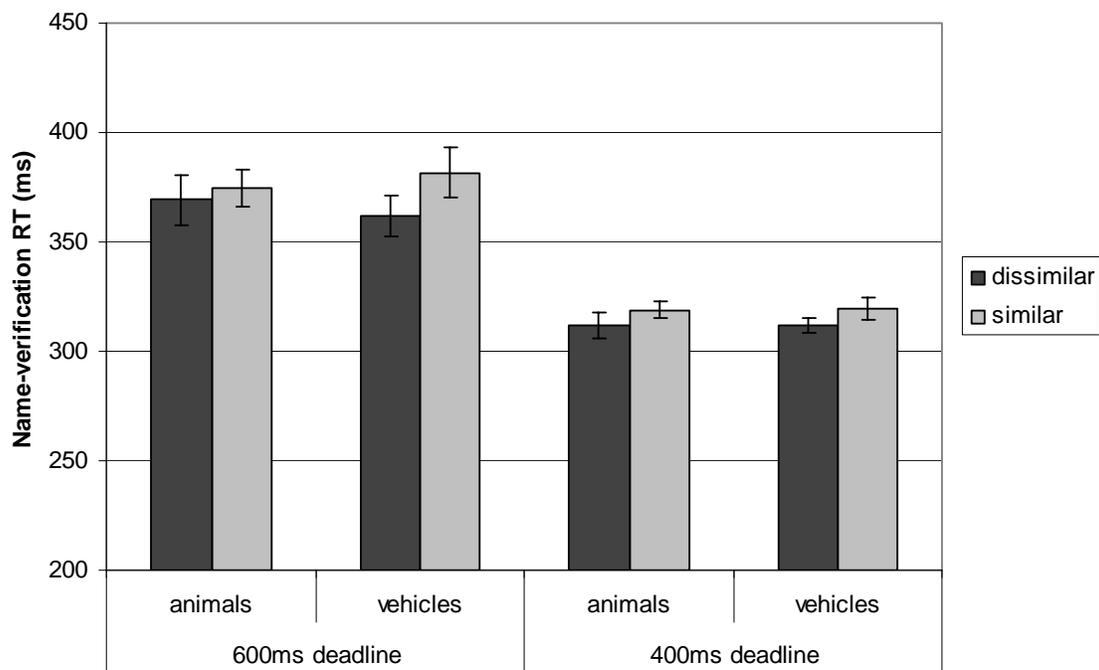


Figure 2

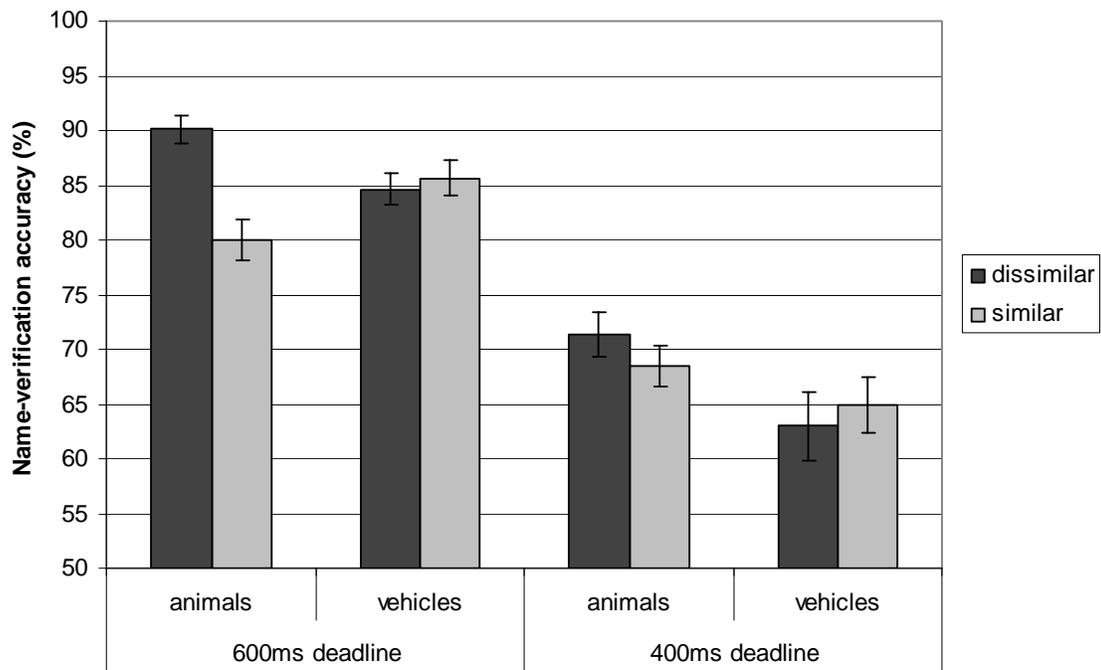


Figure 3

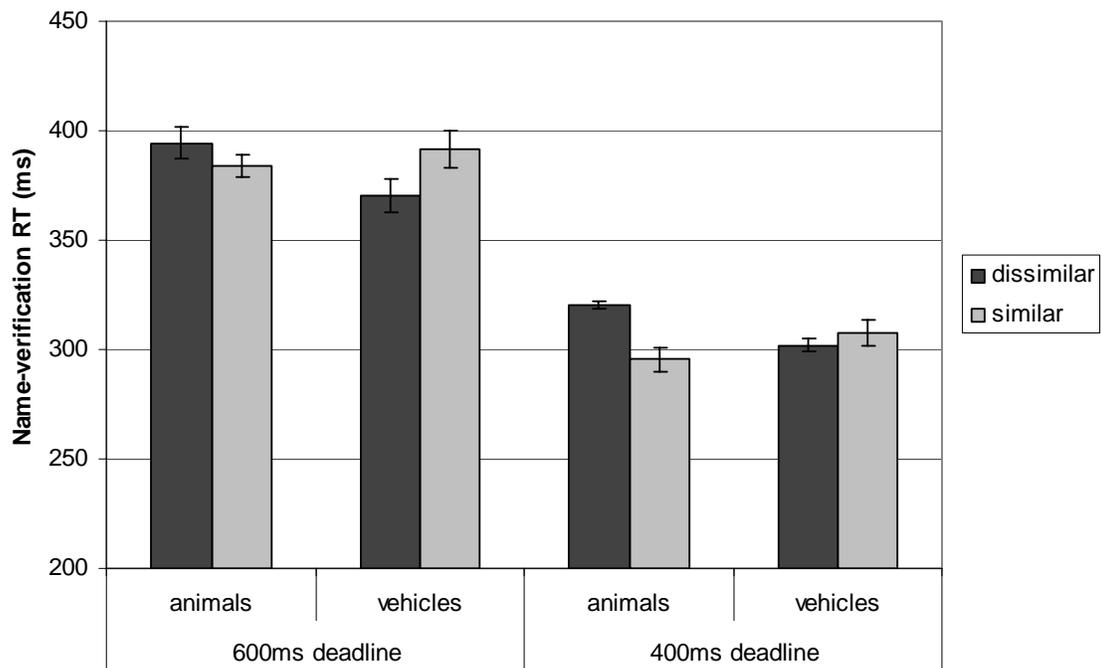


Figure 4