The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

A categorical recall strategy does not explain animacy effects in episodic memory

Joshua E. VanArsdall, James S. Nairne, Josefa N. S. Pandeirada & Mindi Cogdill

To cite this article: Joshua E. VanArsdall, James S. Nairne, Josefa N. S. Pandeirada & Mindi Cogdill (2016): A categorical recall strategy does not explain animacy effects in episodic memory, The Quarterly Journal of Experimental Psychology, DOI: 10.1080/17470218.2016.1159707

To link to this article: http://dx.doi.org/10.1080/17470218.2016.1159707

Accepted author version posted online: 01 Mar 2016.
Published online: 17 Mar 2016.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2016
http://dx.doi.org/10.1080/17470218.2016.1159707

Routledge
Taylor & Francis Group

A categorical recall strategy does not explain animacy effects in episodic memory

Joshua E. VanArsdall[a], James S. Nairne[a], Josefa N. S. Pandeirada[a,b] and Mindi Cogdill[a]

[a]Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA; [b]CINTESIS, Department of Education and Psychology, University of Aveiro, Aveiro, 3810-193, Portugal

ABSTRACT
Animate stimuli are better remembered than matched inanimate stimuli in free recall. Three experiments tested the hypothesis that animacy advantages are due to a more efficient use of a categorical retrieval cue. Experiment 1 developed an “embedded list” procedure that was designed to disrupt participants’ ability to perceive category structure at encoding; a strong animacy effect remained. Experiments 2 and 3 employed animate and inanimate word lists consisting of tightly constrained categories (four-footed animals and furniture). Experiment 2 failed to find an animacy advantage when the categorical structure was readily apparent, but the advantage returned in Experiment 3 when the embedded list procedure was employed using the same target words. These results provide strong evidence against an organizational account of the animacy effect, indicating that the animacy effect in episodic memory is probably due to item-specific factors related to animacy.

ARTICLE HISTORY
Received 19 November 2015
Accepted 24 February 2016

KEYWORDS
Animacy; organizational recall strategies; episodic memory; categories

Converging evidence from several domains suggests that the difference between animate beings and inanimate objects is highly salient, and animates in many cases may enjoy “special” processing advantages over inanimates. Patients’ selective deficits in semantic memory for animals (see Capitani, Laiacona, Mahon, & Caramazza, 2003, for a review), evidence for distinct neural systems dedicated to the detection of agents and animates (Gobbini et al., 2011), and visual systems that respond to animacy as a graded dimension (Sha et al., 2015) suggest that the animate-inanimate distinction is fundamental. Further, animate stimuli appear to afford specialized or preferential processing in many cases. Strong support for the importance of animacy as a dimension comes from developmental research, where it has been shown that biological motion is recognized readily from a very young age (Simion, Regolin, & Bulf, 2008), and the animate–inanimate distinction (roughly described as the difference between living things and non-living things) is one of the first distinctions made by infants and young children (Opfer & Gelman, 2011; Poulin-Dubois, Lepage, & Ferland, 1996).

Research in adults corroborates the importance of animates, indicating that animate stimuli probably attract more attention and are more likely to be noticed—an idea known as the *animate monitoring hypothesis* (New, Cosmides, & Tooby, 2007; Yorzinski, Penkunas, Platt, & Coss, 2014). Recently, it has been reported that animacy affects episodic memory as well: Whether a stimulus such as a word, nonword, or picture is perceived as animate or not influences its later recall on several types of episodic memory tasks. These include free recall (Bonin, Gelin, & Bugaiska, 2014; Bonin, Gelin, Laroche, Méot, & Bugaiska, 2015; Gelin, Bugaiska, Méot, & Bonin, 2015; Nairne, VanArsdall, Pandeirada, Cogdill, & LeBreton, 2013; Popp & Serra, 2016; VanArsdall, Nairne, Pandeirada, & Blunt, 2013), recognition memory (Bonin et al., 2014; VanArsdall et al., 2013), and some types of cued recall (Popp & Serra, 2016; VanArsdall, Nairne, Pandeirada, & Cogdill, 2015).

CONTACT James S. Nairne nairne@psych.purdue.edu Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081, USA

Animate stimuli are consistently remembered better than inanimate stimuli, but a relatively uninteresting factor might explain the advantage. It is possible that animate stimuli simply represent a stronger or more cohesive category than inanimate stimuli—that is, the category "living things" may be more easily noticed or be more diagnostic than the category "non-living things". Researchers have long known that the organizational structure of stimuli can significantly impact retention (e.g., Hunt & Einstein, 1981). Categorized lists are typically remembered better than lists of unrelated words because participants can use the category labels as retrieval cues to help generate possible recall candidates. It is conceivable that participants could readily rely on the category "living things" to help cue their memories for animate stimuli, especially if the "living things" category is more salient or diagnostic than the category of "non-living things".

We refer to this explanation of the animacy effect in episodic memory as the *categorical hypothesis*. The goal of this paper is to examine this hypothesis and to determine whether it explains the animacy effect in free recall. One prediction of the categorical hypothesis is that disrupting a participant's tendency to notice or rely on category information should eliminate or reverse the animacy effect. A potential way of testing this prediction is to use retention tasks in which a categorical recall strategy is less effective. For example, in a recognition memory task the problem is not to remember the item per se, but to remember *whether* the item occurred. A category cue would be useful in helping participants to generate potential items that were seen, but not necessarily helpful in determining which members of the category actually occurred on the memory list. Cued recall is another kind of retention task in which a categorical recall strategy is less likely to be helpful: The problem here is not simply to remember the item, but also to remember its associated cue. Again, a category cue might help the participant generate possible target candidates, but there is no reason to expect it to help one properly pair a target item with its associated cue.

Some research is available on the animacy effect in these types of retrieval tasks. Bonin et al. (2014) have demonstrated the animacy effect in recognition with words, and VanArsdall et al. (2013) have shown animacy advantages in recognition when nonwords are *processed* as animate. The data in cued recall are less clear, however. Whereas VanArsdall et al. (2015) found animacy effects in cued recall tasks resembling foreign-language learning (learning *adui*–duck is easier than learning *adui*–violin, for example), Popp and Serra (2016) have shown that the animacy effect may not be widely generalizable to cued recall. In a number of other types of cued recall tasks (animate cue–animate target, object cue–animate target, etc.), animate targets were recalled the same as or worse than object targets in similar encoding situations.

At this point, no study has attempted to disrupt the use of a categorical retrieval strategy in a free recall environment. Most previous research in free recall has attempted to control for the categorical nature of the lists by equating for variables such as category typicality and size (Van Overschelde, Rawson, & Dunlosky, 2004), as well as semantic relatedness (using latent semantic analysis; see Landauer, Foltz, & Laham, 1998). Although equating these values may not prevent a categorical retrieval strategy, previous results have shown that participants do not cluster their free recall along the animacy dimension (Nairne et al., 2013). The current Experiment 1 takes a different approach. In addition to equating the to-be-remembered words, we also changed the encoding environment to lower the chances that a categorical strategy would be used during recall. To do this, we embedded matched animate and inanimate target words within a larger set of filler words in an effort to mask category structure. As a first attempt with this paradigm, we used words that are known to elicit the animacy effect in free recall—those used by Nairne et al. (2013) in their original demonstration of the animacy advantage.

Experiment 1

In Study 2 of Nairne et al. (2013), participants were asked to remember lists containing an equal number of animate and inanimate words mixed within a list. The animate and inanimate stimuli were equated along a number of mnemonically relevant dimensions (e.g., concreteness, word frequency, relatedness, etc.), including category size and typicality. Still, people may have noticed that half of the words in the list were "living things", and half were not. The goal of Experiment 1 was to replicate the animacy effect found by Nairne et al. (2013) in free recall, using the same matched animate and inanimate words, but in a list context designed to reduce the chances that participants would focus on category structure. Specifically, we embedded only a subset of the matched animate and inanimate target words within a larger set of unrelated filler words. Participants performed three study–test trials in which they received a list of 24 words that contained a random selection of four target animate words, four target inanimate words, 12 filler words, and four "buffer" words, two at the beginning and end of the list. Because the filler words were randomly selected, and differed for each of the three trials, it seemed unlikely that participants would notice and adopt any particular categorical strategy, as they might have in the Nairne et al. study, which employed equal numbers of animate and inanimate words in each list. Again, within a list, only four target animate and inanimate words were presented, but across the three lists, participants were asked to study and recall all 12 of the target animate and inanimate stimuli used in Nairne et al. (2013).

Method

Participants and apparatus

Fifty-five Purdue University undergraduates (29 women) participated in exchange for partial credit in an introductory psychology course. Four participants (one woman) were eliminated for poor performance on at least one of the distractor tasks (more details on this criterion are presented below). Participants were tested in groups of up to four in sessions lasting approximately 30 min. Stimuli were presented and controlled by personal computers, and responses were entered using the keyboard and on recall sheets.

Materials and design

The 24 target words were taken from Nairne et al. (2013). These words included 12 animate and 12 inanimate words matched along 10 mnemonically relevant dimensions; these words have been previously shown to elicit the animacy effect in free recall (for details, see Nairne et al., 2013). A further set of 925 words, taken from Rubin and Friendly's (1986) set of recallability norms (taken from the Paivio, Yuille, & Madigan, 1968, norms), were used to select the random filler words. These words had been previously rated for animacy by three independent raters as reported by Nairne et al. (2013). Of the 925 words, approximately 17% were clearly animate words so the majority of the filler items were unrelated inanimate or ambiguous stimuli (e.g., "*comforter*", "*menace*"). Five words were not used because they overlapped with target words, resulting in a final pool of 920 possible filler words. The experiment was a 2 × 3 repeated measures design with target word type (animate or inanimate) and study–test trial (1, 2, or 3) manipulated within-subject. Number of target words recalled was the dependent variable, and recall order was used to calculate an adjusted-ratio-of-clustering (ARC) score for each participant. Extra-list intrusion data are also reported.

Procedure

People were told that they were participating in a memory experiment and were asked to try to remember each word as it was presented. Each word appeared in the centre of the screen for 5 s with a 250-ms inter-item interval. Each of the three study–test trials contained a unique list of 24 words for participants to study. Each list was composed of four target animate words, four target inanimate words, and 12 filler words (two additional filler words were presented as buffer items at the beginning and end of each list). The words presented in each list were randomly selected without replacement from their respective word pools (target animates, target inanimates, and filler words). The order of presentation of the words in the list was also randomly determined, with the constraint that one target animate and one target inanimate item was presented in each quarter of the list.

Following the final item in each list, to clear working memory, participants completed a short one-minute distractor task in which they made timed decisions about whether a single-digit non-zero number was even or odd; items were presented one at a time on the computer screen for 2 s with an inter-digit interval of 100 ms. Participants who got fewer than 60% of these even–odd judgments correct (performance near or below chance) on any trial were eliminated. The distractor task was followed by 4 min of free recall. For this task, participants were instructed to recall the just-studied list of words in any order. This procedure was repeated three times, one for each list. A different recall sheet was used in each recall task. Except for the manipulation of list composition described above and instructions that made it clear to participants to recall the *most recent* list of words during each recall trial, all aspects of the design, including timing, were held constant across participants and were identical to those used in Study 2 of Nairne et al. (2013).

Results and discussion

Figure 1 shows proportion correct recall of animate and inanimate target words for each of the three study–test trials. For the target words, an effect of animacy occurred on each recall trial. A 2 × 3 repeated measures analysis of variance (ANOVA) confirmed a significant main effect of word type, $F(1, 50) = 39.54$, $MSE = .052$, $\eta_p^2 = 442$, $p < .001$, indicating that a significantly higher proportion of animate targets than of inanimate targets was recalled. The main effect of recall trial was also significant, showing an improvement in memory performance from Trial 1 to Trial 3, perhaps due to practice, $F(2, 100) = 4.60$, $MSE = .068$, $\eta_p^2 = .084$, $p < .05$. The interaction was not statistically significant, $F(2, 100) < 1$.

We also examined recall of filler words and extra-list intrusions during recall. Filler words were recalled at roughly the same rate as inanimate target words (Trial 1 = .37, *SD* = .20; Trial 2 = .44, *SD* = .19; Trial 3 = .46, *SD* = .23). We also looked at intrusion data because intrusions can be indicative of a categorical retrieval strategy. For example, if a greater number of animates are recalled because the animate category is particularly salient, then one might expect to find more animate than inanimate intrusions. Across the three recall trials the average number of intrusions per participant (including extra-list and previous list intrusions) was very low ($M = 2.18$, $SD = 2.15$). Extra-list intrusions—that is, intrusions that were not from prior lists—were classified as animate or inanimate by two of the authors, and, when disagreement occurred, a third researcher was consulted to reach an agreement; those intrusions that could not be clearly classified in one of these groups (e.g., *own*, *illustrate*) were not considered. Results revealed a significantly higher number of inanimate than animate intrusions ($M_A = 0.33$, $SD = 0.622$; $M_I = 1.02$, $SD = 1.64$, respectively), $t(50) = 2.84$, $p < .01$, $d = 0.440$. Thus, the intrusion data provide no obvious support for an animate-based categorical retrieval strategy.

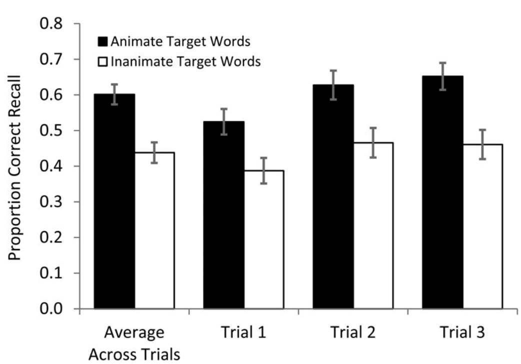

Figure 1. Results from Experiment 1: Mean proportion of target words correctly recalled as a function of trial and word type. Data are shown collapsed cross the three free recall trials and separately for each trial. Error bars represent standard errors of the mean.

A more direct test of category-based retrieval strategies, at least along the animate-inanimate dimension, is to assess the degree of category clustering for each participant. Of particular interest was whether participants would tend to recall by word type (that is, recalling animates together and inanimates together). ARC scores (Roenker, Thompson, & Brown, 1971) were used to estimate this likelihood: An ARC score of zero indicates chance-level clustering, whereas a score of 1.00 indicates perfect clustering. Two participants had to be eliminated from this analysis because it was not possible to calculate ARC scores on every one of their recall trials; this left $N = 49$. Table 1 shows the average ARC scores across all three trials. Random filler words (including buffer words) were included in the calculation of ARC scores (using Nairne et al.'s, 2013, classifications for these words as animate or inanimate) to maximize the chance of observing clustering; intrusions were ignored here. All of the average ARC scores were near zero, indicating that chance clustering occurred. One-sample *t* tests confirmed that none of the average ARC scores was different from zero (all $ts < 1$), and no significant increase in clustering occurred across recall trials, $F(2, 96) < 1$. These data suggest that a categorical recall strategy was not used by participants.

Experiment 2

The data obtained in Experiment 1 are inconsistent with the *categorical hypothesis*. A strong animacy advantage occurred in the absence of any apparent categorical retrieval strategy. Yet there is another way to test the effectiveness of a categorical strategy in free recall—sample both the animate and inanimate target words from tightly constrained and obvious categories. Under these conditions, we reasoned, category retrieval cues should be equally salient for both the animate and inanimate targets, removing any inherent categorical advantage that might have favoured animate targets in previous research. In Experiment 2 we used word lists that were drawn from two highly familiar categories: Four-footed animals and articles of furniture. These stimuli have previously been shown to produce an animacy effect by VanArsdall et al. (2015), although they were only tested in a procedure similar to foreign-language learning.

Table 1. Average ARC scores and standard deviations across the three free recall trials, and separately for each trial.

Experiment	Average across trials		Trial 1		Trial 2		Trial 3	
Experiment 1	−.01	(.218)	−.006	(.387)	−.043	(.499)	.020	(.412)
Experiment 2	.391***	(.265)	.221**	(.446)	.366***	(.335)	.587***	(.303)
Experiment 3	.125*	(.338)	.160	(.576)	.119*	(.386)	.094	(.534)

Note: Standard deviations in parentheses. ARC = adjusted-ratio-of-clustering. These data are presented for each of the experiments. *$p < .05$ for a one-sample t test different from zero. **$p < .01$. ***$p < .001$.

Method

Participants and apparatus

Fifty-four Purdue University undergraduates (26 women) participated in exchange for partial credit in an introductory psychology course. One female participant was eliminated due to poor performance on at least one of the distractor tasks, following the same criteria as those in Experiment 1. Participants were tested in groups of up to four in sessions lasting approximately 30 min. Stimuli were presented and controlled by personal computers, and responses were entered using the keyboard and on recall sheets.

Materials and design

Twenty words from VanArsdall et al. (2015) were used, of which 10 were animate, and 10 were inanimate (see Appendix). These words were chosen because all of the animate and inanimate words belonged to a specific category from the Van Overschelde et al. (2004) category norms ("four-footed animals" and "articles of furniture"). Importantly, VanArsdall et al. also matched these words along a number of potentially important mnemonic dimensions (including concreteness, familiarity, and meaningfulness, among others) in an effort to ensure that animacy was the primary difference between the two sets of words. An additional four words from the same two categories (two animate and two inanimate) were used as "buffer" words. The experiment was a 2 × 3 repeated measures design with target word type (animate or inanimate) and study–test trial (1, 2, or 3) manipulated within subject. Number of words recalled was the dependent variable, and recall order was used to calculate ARC scores for each participant. The nature of the intrusions was also considered.

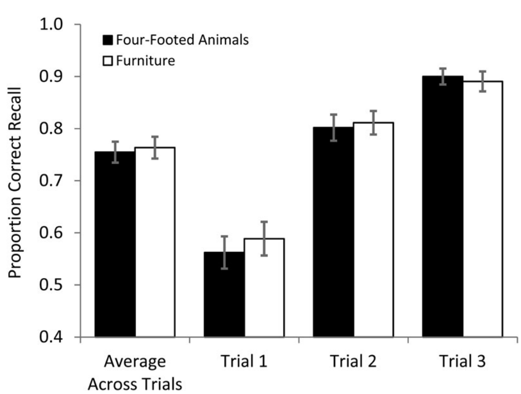

Figure 2. Results from Experiment 2: Mean proportion of words correctly recalled as a function of trial and word type. Data are shown collapsed across the three free recall trials and separately for each trial. Error bars represent standard errors of the mean.

Procedure

Experiment 2 was a direct replication of Study 2 in Nairne et al. (2013), using the highly categorized word lists described above. All 20 words were presented in a different random order to each participant, except that each half of the presented list contained an equal number of animate and inanimate words. Two "buffer" words (one of each type) were presented at the beginning and end of the list; these words were not scored in recall. After studying each list, participants completed a short one-minute distractor task followed by four minutes of free recall. Unlike Experiment 1, each study trial presented the same words, except that words were presented in a new random order on each of the three trials. Except for the words used, all aspects of the design, including timing, were held constant across participants and are identical to those found in Experiment 1.

Results and discussion

The results of the recall task are shown in Figure 2. No recall advantage was present for the animate items on any of the three trials. A 2 × 3 repeated measures ANOVA confirmed that there was no effect of word type, $F(1, 53) < 1$, but an effect of trial was found, $F(1.62, 85.57) = 138.4$, $MSE = .026$, $\eta_p^2 = .723$, $p < .001$; the interaction was not statistically significant, $F(2, 106) < 1$. The number of intrusions was again low, with participants averaging fewer than one intrusion across all three recall trials. In addition, using the classification procedure described in the previous experiment, no differences were found in the number of animate and inanimate intrusions, $t(52) < 1$.

We once again investigated ARC scores to determine whether participants recalled words in a categorical fashion. As before, buffer words were included in the calculation of ARC scores (as their retrieval could still indicate the use of a categorical recall strategy), whereas extra-list intrusions were ignored. Eight of the 53 participants were excluded from the analysis because their recall records did not allow the identification of the order of recall. Interestingly, however, seven of these eight participants were excluded because on at least one trial, they categorized their recalled words into two distinct columns on the recall sheet: one for animate words and one for inanimate words. Although an ARC score could not be calculated for these participants (because it is unclear the order in which words were written into each column), it is very evident that they were using a categorical strategy to aid in their recall of the words. The recall patterns reported above (no effect of word type or interaction, but an effect of recall trial) were true also for the reduced $N = 45$ sample.

Average ARC scores for the 45 participants are shown in Table 1: Each ARC score was significantly different from zero ($p < .01$ in all cases). Once again, ARC scores above zero indicate that greater than chance clustering by category is occurring; these data are very unlike those of Experiment 1. Further, a one-way ANOVA with recall trial as an independent variable revealed that ARC scores increased as a function of trial, $F(1.7, 73.8) = 15.92$, $MSE = .115$, $\eta_p^2 = .266$, $p < .001$, indicating that participants relied more on category structure to aid their recall as the experiment went on.

At first glance, these data seem to support the categorical hypothesis: When a clear categorical recall strategy is available for both animate and inanimate words, no animacy effect is found. According to the hypothesis, the failure to find an animacy effect in this case is presumably due to a boosting of inanimate recall, given that the category "living things" was assumed to be already salient. However, three interpretive problems remain: First, the failure to find any evidence of clustering in Experiment 1, or in Study 2 of Nairne et al. (2013), argues against the idea that the animacy effect depends on participants' use of a categorical retrieval strategy. In fact, those data suggest that animacy effects emerge only when people *fail* to use category cues as part of their retrieval strategy. Second, when people do use a categorical retrieval strategy, as in Experiment 2, there is no easy way to assess the relative effectiveness of each category cue. For example, it is conceivable that the category cue "furniture" is a more effective or more diagnostic retrieval cue than "four-footed animals". If so, then any inherent animacy advantage could be masked by the use of a superior category cue. A parallel study by Gelin et al. (2015) may give some insight into this issue. Gelin et al. demonstrated that when eight categories of words are used (four animate categories and four inanimate categories, with four words each), an animacy effect remains in free recall. Participants may have used a categorical recall strategy in this instance (Gelin et al., 2015, did not present ARC scores), but the animacy advantage was less likely to be masked, as multiple categories result in multiple category cues of varying strengths for both animate and inanimate words. Third, perhaps there is something about these particular word pools, rather than the presence of category clustering, that is responsible for the null effect of animacy. Experiment 3 was designed to test these ideas.

Experiment 3

If the null effect of animacy in Experiment 2 is due to the adoption of a category-based retrieval strategy, rather than to the use of any particular set of animate and inanimate words, then the animacy effect should be restored if we use the embedded list technique of Experiment 1. Embedding a subset of the target words from Experiment 2 into a larger list of randomly selected filler words should mask the categorical nature of the stimuli and reduce the chances that participants will use category-based retrieval cues to drive recall. Participants studied and recalled three lists of 25 words containing a random selection of three target animate words, three target inanimate words, 15 filler words, and four "buffer" words, two at the beginning and end of the list. Each list was different, and, across all three lists, participants saw nine animate (from the category "four-footed animals") and nine inanimate (from the category "an article of furniture") target words from Experiment 2.

Method

Participants and apparatus

Fifty-four Purdue University undergraduates (31 women) participated in exchange for partial credit in an introductory psychology course. Four participants (two women) were eliminated due to poor performance on at least one of the distractor tasks. Participants were tested in groups of up to four in sessions lasting approximately 30 min. Stimuli were presented and controlled by personal computers, and participants entered their responses using the keyboard and on recall sheets.

Materials and design

Eighteen target words were used—nine four-footed animals and nine pieces of furniture taken from Experiment 2; one word of each type was removed (*sheep* and *desk*) from the target set used in Experiment 2, so that three words of each type could be presented per trial. Their removal did not significantly change the distribution of variables for which these words had been equated (see Appendix). Once again, the experiment was a 2 × 3 repeated measures design with target word type (animate or inanimate) and study–test trial (1, 2, or 3) manipulated within subject. Number of target words recalled was the dependent variable, and recall order was used to calculate ARC scores for each participant.

Procedure

As in the previous experiments, participants performed three study–test trials. Each of the lists contained a unique set of words composed of three target animate words, three target inanimate words, and 15 filler words (two additional filler words were presented as buffer items at the beginning and ending of each list; see Experiment 1 for details on filler words), for a total of 25 unique words per list. The words presented in each list were randomly selected without replacement from their respective word pools (target animates, target inanimates, and filler words). Apart from the target words and their distribution across trials, all aspects of the procedure were identical to those in Experiment 1, except the recall period was shortened to three minutes; this was still ample time for participants to complete the recall task.

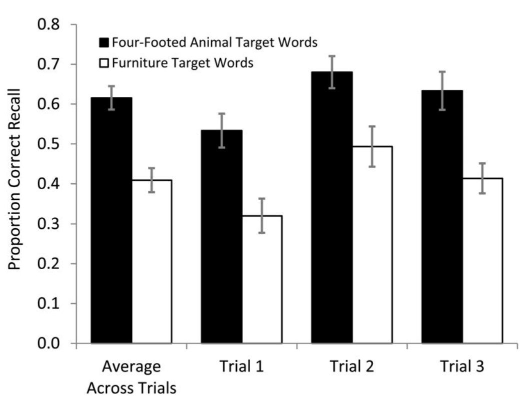

Figure 3. Results from Experiment 3: Mean proportion of target words correctly recalled as a function of trial and word type. Data are shown collapsed across the three free recall trials and separately for each trial. Error bars represent standard errors of the mean.

Results and discussion

Figure 3 shows the results of the recall task. Notably, the observed data are similar to those of Experiment 1. There was a strong animacy effect, which was confirmed by a 2 × 3 repeated measures ANOVA with recall trial and word type as within-subject factors. Both the effects of word type, $F(1, 49) = 49.85$, $MSE = .064$, $\eta_p^2 = .504$, $p < .001$, and recall trial, $F(2, 98) = 8.90$, $MSE = .073$, $\eta_p^2 = .154$, $p < .001$, were significant, whereas the interaction was not, $F(2, 98) < 1$. An animacy effect was present on all three trials, and participants generally improved from Trial 1 to Trial 3. Filler words produced recall levels that were generally closer to the inanimate target words than to the animate target words (Trial 1 = .33, $SD = .15$; Trial 2 = .41, $SD = .15$; Trial 3 = .38, $SD = .16$).

The average number of intrusions across trials in this experiment was similar to that in Experiment 1, averaging about two intrusions per participant ($M = 2.24$, $SD = 3.35$). As before, whenever possible, the intrusions were classified into animates and inanimates following the procedure described in Experiment 1. The average number of inanimate intrusions was significantly higher than the average number of animate intrusions ($M_A = 0.44$, $SD = 1.62$; $M_I = 1.14$, $SD = 1.32$, respectively), $t(49) = 2.58$, $p < .05$, $d = 0.367$.

Average ARC scores were calculated as in Experiment 1 and are shown by recall trial in Table 1. ARC scores could not be calculated for five of the participants and were not considered in this analysis. Little evidence for categorical clustering in recall was found in Experiment 3, although the ARC scores overall were significantly above chance. For the individual recall trials, the ARC value obtained on the second trial differed reliably from chance, but no significant clustering occurred on either the first or the third trials. Of course, strong animacy advantages occurred in recall on every trial in the present experiment. It is important to emphasize that exactly the same target words were used in Experiments 2 and 3, yet whether the animacy advantage appeared (the present experiment) or disappeared (Experiment 2) seemed to depend on whether or not a category-based retrieval strategy was employed during recall.

General discussion

Our experiments were designed to evaluate the *categorical hypothesis* of the animacy effect, the suggestion that animacy effects in episodic memory may be due to an inherent organizational structure that animate words might provide. Previous work on the animacy effect seemed to indicate that a categorical explanation was unlikely, but further investigation was warranted. Previous researchers have attempted to control for within-list categorical variables such as category size and typicality (e.g., Gelin et al., 2015; Nairne et al., 2013; VanArsdall et al., 2015, but not Popp & Serra, 2016), which is useful but does not necessarily prevent participants from using a categorical retrieval strategy. On the other hand, animacy has also been shown by Nairne et al. (2013) to be an important predictor of recall in a regression model using data from recallability norms collected by Rubin and Friendly (1986). These norms were based on the recall of hundreds of words presented in randomly selected lists; thus, it is extremely unlikely that participants focused on animacy as a category and employed animacy-based retrieval cues during recall in these experiments. Animacy effects have also been demonstrated in retrieval environments in which a categorical recall strategy is less useful, such as recognition and cued recall (Bonin et al., 2014; VanArsdall et al., 2013, 2015), but those studies say little about the processes used during free recall.

The present experiments directly manipulated the salience of category structure during encoding, either to eliminate it (Experiments 1 and 3) or to make it obvious (Experiment 2). Across experiments, our results were largely inconsistent with the categorical explanation. In Experiment 2, when animate and inanimate categories were obvious, and significant clustering occurred, the animacy advantage disappeared. By embedding target animate and inanimate words within a larger list of randomly selected words, we were able to largely disrupt participants' use of category structure as a means to guide recall, as evidenced by mostly chance-level ARC scores (Experiments 1 and 3). In both of the experiments using this embedded list technique there was little evidence of category clustering, yet strong and consistent animacy advantages were obtained. Extra-list intrusions can be diagnostic as well because they can indicate whether participants relied on any form of categorical or schematic processing when recalling information. Using categorical strategies usually results in a higher number of intrusions (Reyna & Brainerd, 1995; Roediger & McDermott, 1995). But participants generally produced more inanimate than animate intrusions, suggesting that recall was not especially driven by the use of an animate category. Further, our reported intrusion data are consistent with those found in both Bonin et al. (2015) and Gelin et al. (2015). Collectively, these results argue strongly against the categorical hypothesis—if anything, relying on a categorical recall strategy eliminates the animacy advantage rather than produces it.

The present experiments also demonstrate, yet again, the powerful effect that animacy can have on retention. Animacy as a dimension was initially investigated because of its likely adaptive significance (Barrett, 2005; Nairne et al., 2013), and it is important to note that even if the proximate mechanism of the animacy effect turned out to be categorical in nature, it would not diminish the finding: The ultimate adaptive goal of noticing and remembering animates could easily have been solved by a proximate mechanism that involved more efficient use of animate categories (see Nairne, 2014, for a review of ultimate and proximate mechanisms). The present data seem to indicate, however, that this is not the case; categorical organization is not reflected in the manner that participants output their recall—although in theory participants could have enhanced access to the "animate" category, but not use it to structure their output. We should also note that the present experiments did not explicitly evaluate the role that the detection of animacy may play at encoding. For example, if people are told to attend to the animate-inanimate dimension at encoding, it is quite possible that the animacy advantage would be enhanced or, in some cases (e.g., Experiment 2), eliminated. Our experiments merely show that in the absence of a categorical retrieval strategy, the animacy advantage remains intact.

Other proximate mechanisms for the animacy effect have been proposed, including an attentional bias for animates (VanArsdall et al., 2013; consistent with New et al., 2007), greater use of interactive imagery (Bonin et al., 2015), or possibly arousal (Popp & Serra, 2016). Animate words might be more distinctive in these list contexts as well. In the Rubin and Friendly (1986) norms, for example, only about 17% of the words are clearly living things. Although our lists contained the same number of animate and inanimate words in Experiment 2 (in which no animacy effect occurred) and in Nairne et al. (2013; in which a robust animacy effect was obtained), inanimate words occurred much more frequently in the embedded lists of Experiments 1 and 3 (comprising approximately 70% of the list items). Thus, the animate words might have been more distinctive, especially in the embedded lists, because they represented a smaller proportion of the total word pool. However, distinctiveness seems unlikely to account generally for animacy advantages for two reasons. First, the failure to detect any evidence of clustering in Experiments 1 and 3, and in Nairne et al., suggests that participants fail to note and use the animacy dimension during recall. Second, animacy effects have been detected in between-list designs, where lists are composed entirely of animate or inanimate words (e.g., Popp & Serra, 2016), and distinctiveness effects in memory tend to occur primarily in within-subject designs.

Proximate accounts of animacy effects in retention need to be developed, but regardless of the proximate mechanism, the powerful effect that animacy can have on retention (as demonstrated here and in prior work) has significant implications for cognitive research. For example, cognitive researchers typically fail to control for animacy in their word sets, and, as Nairne et al. (2013) showed, animacy is an extremely robust predictor of recall. Animacy is at least as strong a predictor of recall as imagery, word frequency, and other variables that are commonly controlled in memory experiments. The animacy dimension is also important because its effects on retention are consistent with the view that memory systems are tuned or biased to variables that potentially impact fitness (see Nairne, 2010; Nairne, Thompson, & Pandeirada, 2007). To understand memory functioning completely, it is helpful to know the kinds of problems that our memory systems evolved to solve. The detection and retention of animates is clearly relevant to fitness, so it is perhaps not surprising that our retention systems show sensitivity to this dimension.

Funding

This research was supported, in part, by grants from the National Science Foundation [grant number BCS-0843165], [grant number BCS-1532345]. Josefa N. S. Pandeirada was supported, in part, by the Portuguese Foundation for Science and Technology Fellowship [grant number IF/00058/2012] and research grant [grant number IF/00058/2012/CP0172/CT0002]. Josefa N. S. Pandeirada is also affiliated to the Center for Health Technology and Services Research (CINTESIS).

References

Barrett, H. C. (2005). Adaptations to predators and prey. In D. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 200–223). Hoboken, NJ: Wileys & Sons.

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition*, 42, 370–382. doi:10.3758/s13421-013-0368-8

Bonin, P., Gelin, M., Laroche, B., Méot, A. & Bugaiska, A. (2015). The "how" of animacy effects in episodic memory. *Experimental Psychology*, 62, 371–384. doi:10.1027/1618-3169/a000308

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, 20, 213–261. doi:10.1080/02643290244000266

Coltheart M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A*, 33, 497–505. doi:10.1080/14640748108400805

Gelin, M., Bugaiska, A., Méot, A., & Bonin, P. (2015). Are animacy effects in episodic memory independent of encoding instructions? *Memory*. Advance online publication. doi:10.1080/09658211.2015.1117643

Gobbini, M. I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., … Pietrini, P. (2011). Distinct neural systems involved in agency and animacy detection. *Journal of Cognitive Neuroscience*, 23, 1911–1920. doi:10.1162/jocn.2010.21574

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, 20, 497–514. doi:10.1016/S0022-5371(81)90138-9

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284. doi:10.1080/01638539809545028

Nairne, J. S. (2010). Adaptive memory: Evolutionary constraints on remembering. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. *53*, pp. 1–32). Burlington, VT: Academic Press. doi:10.1016/S0079-7421(10)53001-9

Nairne, J. S. (2014). Adaptive memory: Controversies and future directions. In B. L. Schwartz, M. L. Howe, M. P. Toglia, & H. Otgaar (Eds.), *What is adaptive about adaptive memory?* (pp. 308–321). New York, NY: Oxford University Press.

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 263–273. doi:10.1037/0278-7393.33.2.263

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, 24, 2099–2105. doi:10.1177/0956797613480803

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, 104, 16598–16603. doi:10.1073/pnas.0703913104

Opfer, J. E., & Gelman, S. A. (2011). Development of the animate-inanimate distinction. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 213–238). Oxford: Wiley-Blackwell.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 76, 1–25. doi:10.1037/h0025327

Popp, E. Y., & Serra, M. J. (2016). Adaptive memory: Animacy enhances free recall but impairs cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 186–201. doi:10.1037/xlm0000174

Poulin-Dubois, D., Lepage, A., & Ferland, D. (1996). Infants' concept of animacy. *Cognitive Development*, 11, 19–36. doi:10.1016/S0885-2014(96)90026-X

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75. doi:10.1016/1041-6080(95)90031-4

Roediger, H. L. III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803–814. doi:10.1037/0278-7393.21.4.803

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 76, 45–48. doi:10.1037/h0031355

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition*, 14, 79–94. doi:10.3758/BF03209231

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience*, 27, 665–678. doi:10.1162/jocn_a_00733

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, 105, 809–813. doi:10.1073/pnas.0707021105

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Blunt, J. R. (2013). Adaptive memory: Animacy processing produces mnemonic advantages. *Experimental Psychology*, 60, 172–178. doi:10.1027/1618-3169/a000186

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Cogdill, M. (2015). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory*, 23, 657–663. doi:10.1080/09658211.2014.916304

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, 50, 289–335. doi:10.1016/j.jml.2003.10.003

Yorzinski, J. L., Penkunas, M. J., Platt, M. L., & Coss, R. C. (2014). Dangerous animals capture and maintain attention in humans. *Evolutionary Psychology*, 12, 534–548. doi:10.1177/147470491401200304

APPENDIX

Table A1. Animate and inanimate English words used in each experiment.

Experiment 1		Experiments 2 & 3	
Animate	Inanimate	Animate	Inanimate
baby	doll	bear	bed
bee	drum	cat	cabinet
duck	hat	fox	chair
engineer	journal	mouse	couch
minister	kite	rabbit	desk[a]
owl	purse	rat	dresser
python	rake	sheep[a]	lamp
soldier	slipper	tiger	sofa
spider	stove	turtle	stool
trout	tent	wolf	table
turtle	violin		
wolf	whistle		

[a]Words not used in Experiment 3.

Table A2. Statistical characteristics of the control variables in Experiments 1–3 for animate and inanimate stimuli.

		Animate			Inanimate			
Experiment	Dimension	Mean	*SD*	Range	Mean	*SD*	Range	*t*
Experiment 1	Category size	22.25	5.91	14–29	23.17	6.03	17–33	$\|t\| < 1$
	Category typicality	0.200	0.165	0.06–0.51	0.238	0.182	0.05–0.62	$\|t\| < 1$
	Concreteness	593.4	28.85	531–644	592	16.77	563–626	$\|t\| < 1$
	Familiarity	503.9	69.69	308–597	506.5	30.62	468–580	$\|t\| < 1$
	Kučera–Francis frequency	21.67	22.84	2–62	16.50	16.36	1–56	$\|t\| < 1$
	Imagery	588.5	37.17	495–632	578	30.39	509–624	$\|t\| < 1$
	Meaningfulness	447.5	55.59	364–573	438.2	32.03	380–487	$\|t\| < 1$
	Number of letters	5.33	1.78	3–8	5	1.41	3–7	$\|t\| < 1$
	Relatedness	0.114	0.120	0–0.55	0.143	0.125	−0.07–0.53	$t = -1.35$, *ns*
Experiment 2	Category size	28	*n/a*	*n/a*	21	*n/a*	*n/a*	*n/a*
	Category typicality	0.259	0.274	0.06–0.97	0.436	0.291	0.06–0.9	$t = -1.40$, *ns*
	Concreteness	616	17.82	585–644	599.5	23.16	560–635	$t = 1.79$, *ns*
	Familiarity	526.6	24.10	501–582	562.7	50.12	472–636	$t = -2.05$, *ns*
	Kučera–Francis frequency	16.40	15.61	6–57	51.80	64.93	1–198	$t = -1.68$, *ns*
	Imagery	598.6	18.38	564–617	577.3	33.12	524–635	$t = 1.78$, *ns*
	Meaningfulness	441.5	22.88	408–487	444.9	53.61	374–545	$\|t\| < 1$
	Number of letters	4.40	1.17	3–6	4.90	1.29	3–7	$\|t\| < 1$
	Relatedness	0.311	0.098	0.121–0.441	0.335	0.114	0.119–0.509	$\|t\| < 1$
Experiment 3	Category size	28	*n/a*	*n/a*	21	*n/a*	*n/a*	*n/a*
	Category typicality	0.277	0.284	0.06–0.97	0.430	0.308	0.06–0.9	$t = -1.10$, *ns*
	Concreteness	615.3	18.77	585–644	601.3	23.78	560–635	$t = 1.39$, *ns*
	Familiarity	528.8	24.49	501–582	560.4	52.61	472–636	$t = -1.64$, *ns*
	Kučera–Francis frequency	15.67	16.37	6–57	50.33	68.69	1–198	$t = -1.47$, *ns*
	Imagery	598.9	19.47	564–617	577.7	35.10	524–635	$t = 1.59$, *ns*
	Meaningfulness	441.2	24.25	408–487	447.9	56.51	374–545	$\|t\| < 1$
	Number of letters	4.33	1.22	3–6	5	1.32	3–7	$t = -1.11$, *ns*
	Relatedness	0.358	0.085	0.209–0.478	0.347	0.125	0.108–0.511	$\|t\| < 1$

Note: Statistical characteristics: means, standard deviations, minimum–maximum range, and independent *t* tests of the means. Values for category size and category typicality were taken from the Van Overschelde et al. (2004) category norms. Values for concreteness, familiarity, imagery, Kučera–Francis written frequency, meaningfulness (Colorado norms), and number of letters were obtained from the MRC Psycholinguistic Database (Coltheart, 1981). The relatedness dimension was assessed using latent semantic analysis (Landauer et al., 1998). A meaningfulness value was not available for the word "cabinet".