Title
It’s Alive! Animate Sources Produce Mnemonic Benefits

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Snoddy, Sean
Silliman, Daniel C.
Wilson, Joseph C.
et al.

Publication Date
2019

Peer reviewed

It's Alive! Animate Sources Produce Mnemonic Benefits

Sean Snoddy (ssnoddy1@binghamton.edu)
Daniel C. Silliman (dsillim1@binghamton.edu)
Joseph C. Wilson (jwilso10@binghamton.edu)
Kenneth. J. Houghton (khought2@binghamton.edu)
Deanne L. Westerman (wester@binghamton.edu)

Department of Psychology, Binghamton University (SUNY)
Binghamton, NY 13902 USA

Abstract

The mnemonic benefits of animate (e.g., Tiger) over inanimate (e.g., Table) stimuli have been demonstrated across several different memory paradigms. Given the ubiquity of inanimate, computer-generated voices we investigated if the animacy of a presentation source confers mnemonic benefits. We asked: is information delivered by a human voice better remembered than information presented by a computer-generated voice? Word-lists were presented auditorily by either a human or a computer-generated voice and memory was measured using a free recall assessment. In Experiment 1, words presented in a human voice were better remembered than words presented in a computer voice. Experiment 2 demonstrated that beliefs about the animacy of a computer-generated voice were not sufficient for any benefits to accrue, suggesting a possible boundary condition for the effect. Both experiments replicated the mnemonic benefits of animate words and demonstrated further extensions of the effect to spoken word presentation.

Keywords: Animacy; Recall; Memory

Introduction

Evolutionary psychologists have long argued that our minds have been adapted through the forces of natural selection (Cosmides & Tooby, 1994). Extending this evolutionary logic, it is further argued that our memory system has been adapted to serve the purposes of surviving in our distant ancestral environments. A recent example of this work would be the effect of "survival processing" by which mnemonic benefits are observed for stimuli experienced in evolutionarily salient contexts (Nairne, Thompson, & Pandeirada, 2007). Another example of adaptive memory is the finding of superior memory for animate compared to inanimate stimuli (Bonin, Gelin, & Bugaiska, 2014; Nairne, VanArsdall, Pandeirada, Cogdill, & LeBreton, 2013; VanArsdall, Nairne, Pandeirada, & Cogdill, 2015).

The *animacy effect* (henceforth *item-animacy*) has been observed in several memory paradigms such as free recall (Nairne et al., 2013), paired-associate recall (VanArsdall et al., 2015), and recognition (Bonin et al., 2014). Nairne and colleagues (2013) posit that our memories would be better attuned to animate entities in the environment for several evolutionary reasons. These include the special threat that living entities can pose, the sustenance that they can provide, and their broad social utility given that interactions with other animate entities (e.g. humans) were crucial for survival and reproduction.

It is this last reason relating to human sociality that drives the current investigation. The central question considered here is: does the animacy of the source of information matter for memory performance? In our modern computer-age, we are constantly interacting with voices generated by computers. How does the perceived humanness of such voices affect our cognition? Could it be that information delivered by Siri would be remembered differently than information provided by an actual person? It's possible that the findings regarding the animacy effect might bear on such questions. To the extent that such computer voices are perceived as inanimate (or at least less animate), there is a possibility that our memories might be worse for the information produced by a computer voice.

This ostensible *source-animacy* effect might emerge due to possible animacy contamination mechanisms (Cogdill, Nairne, & Pandeirada, 2016; as cited in Nairne, VanArsdall, & Cogdill, 2017). For example, Nairne and colleagues (2017) had participants read sentences in which two objects come in contact with each other. Target inanimate words in these sentences are "touched" by either animate ("The mouse is touching the sled.") or inanimate ("The lamp is touching the bottle.") stimuli. They found superior recall performance for inanimate target words when they were "touched" by the animate stimuli as compared to inanimate ones. Nairne and colleagues suggested that the "law of contamination" (Rozin, Millman, & Nemeroff, 1986) may account for such effects, with the property of animacy being conferred contagiously to inanimate words. Therefore, it may be the case that words spoken by the human voice are "contaminated" by the animacy of the voice, thus conferring a benefit for their recall.

Another account points to the importance of the voice itself. The quality of humanness in auditory perception might be especially well-processed. Evidence suggests that, from infancy, there is a predilection for human speech over non-speech analogues (Vouloumanos & Werker, 2007). There is also precedent in the music literature regarding the importance of human vocality for memory. For example, melodies sung by humans are better remembered than instrumental melodies (Weiss, Trehub, & Schellenberg, 2012). The authors proposed that we are especially attuned to human timbres because of their biological significance.

Due to the paucity of research on the subject and the implications for our interactions with machines in daily life, the current study was undertaken. The present experiments employed a free recall test on stimuli delivered through the auditory modality. The central manipulation involved the animacy of the voice delivering the word lists to be recalled (human vs. computer-generated). There were three main objectives: 1) To examine the influence of animate vs. inanimate sources on recall, 2) To provide a direct reproduction of the standard item-animacy effect in an auditory modality with a free recall assessment (see Aslan & John, 2016 for a paired-associate animacy effect using the auditory modality; see Stori, Zaar, Cooke, & Mattys, 2018 for a recognition memory assessment), and 3) To explore whether there would be an interaction between item-animacy and source-animacy. Following the evolutionary reasoning of Nairne and colleagues (2013), superior recall should be evidenced for words delivered by the human (animate) voice.

Experiment 1

The aim of Experiment 1 was to extend the classical item-animacy effect to an auditory source paradigm. Past research exploring animacy has typically consisted of the visual presentation of word lists that included animate and inanimate words (Nairne et al., 2013).The key departure from many past studies is that these lists are presented aurally through two different voices to manipulate source-animacy (cf. Aslan & John, 2016) along with a free recall assessment (cf. Stori et al., 2018). One of these voices was human and the other was computer-generated. Based on the animacy and evolutionary literature, memory should be superior for those words presented by the human-voiced (animate) compared to the computer-voiced (inanimate) source. Furthermore, this paradigm should replicate the standard item-animacy effect.

Method

Participants Binghamton University undergraduates (N = 51) participated in this study. An additional participant did not complete the entire experiment and was excluded.

Materials and Design Thirty-six English words were used in this experiment (18 animate, 18 inanimate). Thirty-two of these words (17 animate, 15 inanimate) were selected from word lists used in Nairne and colleagues (2013) and VanArsdall and colleagues (2015). An additional 1 animate and 3 inanimate words were obtained from the MRC database (Wilson, 1988) to supplement the lists. Following Nairne and colleagues (2013), all words were concrete nouns and matched on several dimensions: age of acquisition (19 words were missing data), number of letters, familiarity, imageability, concreteness, Kučera and Francis written frequency and number of categories, and mean Colorado meaningfulness.

Two versions of each word were recorded using version 2.1.3 of Audacity® (Audacity, 2014). The human spoken words were recorded by an experimenter that read each word aloud into the built-in microphone of an Apple Macbook laptop computer. A second Macbook computer was used to recreate the same set of words voiced by a computer using the voice-over accessibility function that comes standard with Apple computers and recorded via the built-in microphone of the first Macbook. The result of each recording was a continuous WAV file for each human- and computer-voiced word list. These continuous files were edited into discrete WAV files in Audacity for all of the words in both human- and computer-voiced presentations. All words were adjusted to have comparable volumes in both the human- and computer-voiced conditions (range: 9-15 dB). As a pilot test for clarity in the presentation of the words, a research assistant listened to both human- and computer-voiced presentations of all words and wrote them down. All words used in the present study were clearly perceptible to the research assistant, however, an additional four words were unclear and instead used as buffer words.

Two lists of intermixed animate and inanimate words were used for each experiment session and randomly assigned to either human- or computer-voiced conditions. Words were assigned to each list such that both human- and computer-voiced conditions were balanced on the aforementioned item-level variables. Two fixed buffer words were presented at the beginning of the first list and the end of the second list. Recall for these words were not coded nor included in the final analyses. PsychoPy psychophysics software version 1.8.3 (Peirce, 2007) was used to randomly select the word-to-list assignment, and to present each list in a randomly determined order.

Recall packets were printed on paper and included a maze (distractor task), a blank recall sheet, and a four-item questionnaire to assess the clarity and pleasantness of each word list using 7-point Likert scales.

Procedure For each session, between one and five participants were seated in a quiet room and told that they were participating in a memory experiment. The experiment was displayed on a 48 in. LCD television. Participants were presented with instructions both verbally and on-screen. They were instructed to face forward during presentation of the word lists and to focus on a black fixation cross on a white background. The words were not presented visually on the screen, but the display helped ensure that all participants were attending to the list presentation. Additionally, they were instructed to listen carefully and to expect a recall test later in the experiment.

Participants were then presented with the two word lists via speakers on the television. Each list was either presented using the human- or computer-voiced recordings in their entirety, with the alternative list being subsequently presented. Counterbalancing, which occurred across sessions, determined what source they heard first (human- or computer-voiced). The presentation of the first and the second list was separated by a 30 second break. Following the presentation of the second list, participants were handed a recall packet and directed to begin the maze distractor task. After one minute, participants were instructed to flip the page and recall as many words as they could from both word lists. Participants were given an unlimited amount of time for recall but were told that they could turn the page if they could not remember any more words. On the final page of the packet, participants were asked to indicate the level of clarity and pleasantness of each source-animacy condition.

The following criteria were used to score a participant response as a correct recollection: correctly spelled target words (e.g., 'rabbit'); incorrectly spelled, but closely approximated target words (e.g., 'rabit'); different forms (i.e., tense, plurality) of target words (e.g., 'rabbits'). Responses that were confusable with a non-target word (e.g., 'rabid') were not counted as a correct recollection.

Results and Discussion

There were two main predictions about recall. First, that the animacy effect would be replicated with an auditory presentation of word lists and free recall assessment. Specifically, recall performance would be higher for animate words than inanimate words. Second, it was predicted that presentations voiced by a human should lead to better recall than presentations from a computer. A repeated measures ANOVA that contained item-animacy and source-animacy tested these predictions. There was a main effect of item-animacy such that animate words (M = .298, SD = .180) were recalled at a higher rate than inanimate words (M = .192, SD = .159), $F(1, 49) = 13.557$, $p < .001$, $\eta_p^2 = .217$. There was also an effect of source-animacy, where human-voiced words (M = .288, SD = .158) were recalled at a higher rate than computer-voiced words (M = .203, SD = .186), $F(1, 49) = 21.401$, $p < .001$, $\eta_p^2 = .304$. There was no significant interaction between item- and source-animacy ($F < 1$). (See Figure 1).

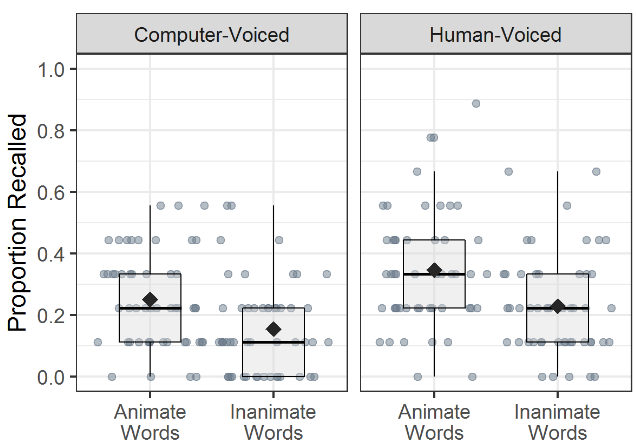

Figure 1: Proportion of words recalled in Experiment 1. The left panel presents both animate and inanimate words in the computer-voiced source animacy condition while the right panel reflects the human-voiced condition. Each point represents a participant's proportion of words recalled. Diamonds represent the overall mean for each condition.

Cumulative-link regression models were used to assess if human-voiced words were perceived as clearer and more pleasant than computer-voiced words. Each model predicted the rating of interest with source-animacy, the presentation order of human- and computer-voiced sources, and their interaction. Human-voiced words were rated as clearer than computer-voiced words (β = 1.873, SE = 0.525, Wald Z = 3.569, $p < .001$). There was no significant difference in ratings based on the order in which human- and computer-voiced sources were presented (β = -0.585, SE = 0.513, Wald Z = -1.140, p = .254) and no significant interaction (β = 0. 432, SE = 0. 713, Wald Z = 0.606, p = .544). Likewise, the human-voiced source received significantly higher pleasantness ratings than the computer-voiced source (β = 2.399, SE = 0.548, Wald Z = 4.378, $p < .001$). Again, there were no significant differences in ratings based on the order in which human- and computer-voiced sources were presented (β = 0.852, SE = 0.537, Wald Z = 1.587, p = .112), and no significant interaction (β = -0.871, SE = 0.727, Wald Z = -1.198, p = .321). (See Figure 2).

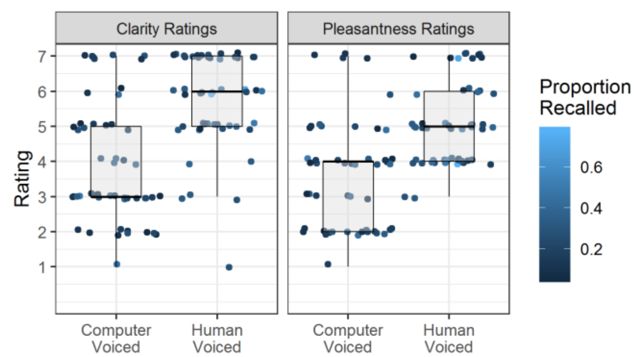

Figure 2: Clarity (left) and pleasantness (right) ratings for each source-animacy condition. Each point represents an individual participants' rating. The shading of each point reflects that participant's proportion of successfully recalled words within each condition. While both clarity and pleasantness ratings differed between source-animacy conditions there was no relationship between ratings and recall.

Mixed-effects logistic regression models that predicted recall success of each word tested if a series of control variables could account for either animacy effect. The baseline model included participants as random intercepts, source-animacy as random slopes, and source-animacy, item-animacy, and their interaction as fixed effects. Control variables were individually entered into the baseline model as an additional fixed effect. Clarity ratings (β = -0.075, SE = 0.050, Wald Z = -1.507, p = .132), pleasantness ratings (β = -0.058, SE = 0.055, Wald Z = -1.07, p = .287), the list participants received (β = 0 .035, SE = 0.171, Wald Z = 0.203, p = .839), and counterbalance order of source-animacy conditions (β = -0.105, SE = 0.167, Wald Z = -0.630, p = .529) were not predictive of recall successes. None of these variables altered the significance of item- and source-animacy effects, or significantly improved the model's ability to account for variance in recall (all *p*s > .14).

Experiment 1 replicated and extended the standard item-animacy effect by demonstrating that animate words were better recalled than inanimate words when using auditorily presented stimuli. Animacy effects were not just observed for items, but also for the sources that presented items. This is reflected by the source-animacy effect: words presented by an animate, human voice were better remembered than items presented by an inanimate, computer-generated voice. The human-voiced source was rated as clearer and more pleasant than the computer-voiced source, however, follow-up analyses revealed that differences in ratings between the presentation source conditions could not account for the source-animacy effect. Taken together, these results provide evidence of the systemic effects of animacy on human memory.

Experiment 2

A potential limitation of the previous experiment was that there may have been differences between the human and computer voice that were not controlled for and that are not related to animacy or evolutionary mechanisms. One such difference was the human-voiced source being rated as clearer than the computer-voiced source (although clarity was not found to be predictive of recall success). The difficulty in controlling human and computer voices across relevant dimensions such as familiarity, tonality, and articulation (which may all contribute to clarity) raised the question of whether the source-animacy effect is contingent upon these differences (i.e. it is due to intrinsic qualities of the human voice) or participants' beliefs about the animacy of the source. To address this question, Experiment 2 circumvented the issue of auditory differences entirely. Instead of two different voices, the words in Experiment 2 are all delivered by one computer-generated voice. While the source (i.e. the voice) was held constant for both conditions, the belief regarding the animacy of the source was manipulated between conditions. Those in the stated-computer condition were told that the voice is computer-generated, while those in the stated-human condition were told that the voice is human.

Instead of serving as a direct replication of Experiment 1, the present experiment tested two hypotheses about the source-animacy effect. The belief-based hypothesis states that the source-animacy effect is determined by participants' belief about animacy independent of the auditory signal. This hypothesis predicts that when participants are presented with a computer-voiced source and their belief in the animacy of the source is manipulated, a source-animacy effect will be observed between the stated-human and stated-computer conditions. The intrinsic qualities hypothesis states that the source-animacy effect is determined by intrinsic qualities of the source. This hypothesis predicts that when presented with a computer-voiced source, no differences between the stated-human and stated-computer conditions will emerge as they are listening to the same computer-generated auditory signal.

Method

Participants Binghamton University undergraduates (*N* = 95) participated in this experiment. Two additional participants were dropped due to technical problems.

Materials and Design The word stimuli were the same as those used in Experiment 1 except that the buffer words were omitted. The audio was produced using Natural Readers online software, a text-to-speech tool[1]. All words recorded for this experiment were produced using a single computer voice from this software, which resembled a British-accented male. The procedure used to convert each word into an audio file was identical to Experiment 1, except that a human voice was not also recorded.

The WAV files for each word were presented through PsychoPy software in a random order to each participant. A between-subjects presentation of the words was used, such that all 36 words were presented to each participant through the one computer voice—the only difference being whether the participant was told that the voice was human or a computer program. In this way, all participants heard the same audio, ensuring that there were no aural or linguistic confounds between the animate and inanimate conditions.

Procedure Participants were randomly assigned to either the stated-computer or stated-human condition. In the stated-computer condition, the participants were told that each word was produced by a computer and in the stated-human condition, they were told that the words were produced by a human.

Each participant was brought individually into a room and told that they would be participating in an experiment that would require them to judge the clarity of a series of words that were to be used as part of a later experiment, which they would *not* be participating in on that day. These clarity judgments served as an incidental encoding task that was followed by a surprise free recall test that immediately followed the clarity judgment task. Participants were provided with closed ear headphones to listen to the words.

Each word of the study list was presented aurally through the headphones in a randomized order across the 36 trials. During each trial, a fixation cross appeared on the screen to focus their attention while the words were presented. A clarity rating scale replaced the fixation cross at the onset of each word. The participant would render their clarity rating on a 5-point Likert scale, with the wording being different according to the condition they were in ("Please rate how clear this human/computer produced word is", 1= not at all clear, 5 = extremely clear). Selecting a rating on the scale would begin the next trial (i.e. the following word).

[1] Navigate to https://www.naturalreaders.com/online/ to access the text-to-speech tool. The voice used was Peter at -1 speed.

At the end of the clarity judgment phase, participants were asked to recall as many words as possible from the list they just heard. They were given an unlimited amount of time to type their responses into an array of boxes that appeared on the screen. Once they completed this recall session, participants were asked how much they believed in the story they were told in the beginning of the experiment as a manipulation check. Those in the stated-computer condition were asked the extent to which they believed the voice they heard came from a computer, while those in the stated-human condition were asked how much they believed the voice to be from a human. Participants were probed about their beliefs on a 5-point scale. The criteria for a correct recollection were the same as in Experiment 1.

Results and Discussion

Data were first analyzed using a two-way ANOVA, with item-animacy as a within-subjects factor and source-animacy as a between-subjects factor. In line with our predictions, the standard animacy effect was replicated in this analysis as a main effect for item-animacy, $F(1, 93) = 52.008$, $p < .001$, $\eta_p^2 = .359$, with a greater proportion of animate words ($M = 0.25$, $SD = .10$) recalled than inanimate words ($M = 0.16$, $SD = .10$). However, there was no main effect found for source-animacy ($F < 1$). No interaction was found between item- and source-animacy ($F < 1$). (See Figure 3). Given the lack of an effect of source-animacy and the intrinsic qualities hypothesis's prediction of a null result, the stated-computer ($M = .202$, $SD =.129$) and stated-human ($M = .216$, $SD =.107$) conditions were analyzed with a Bayesian independent samples t-test. The Bayes Factor indicated substantial support (Jefferies, 1961) for the null hypothesis, (i.e., no differences between conditions), $BF_{01} = 4.425$.

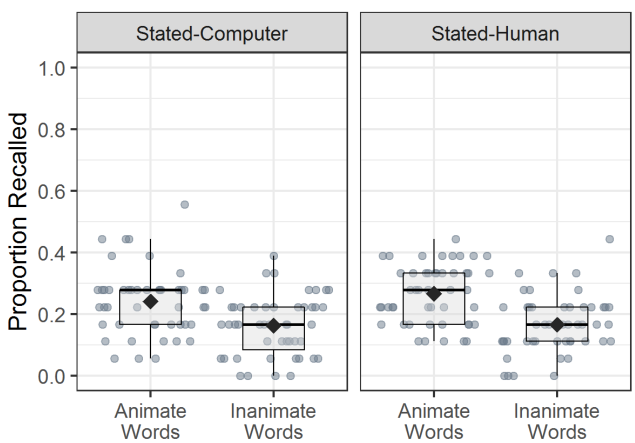

Figure 3: Proportion of words recalled in Experiment 2. The left panel presents both animate and inanimate words in the stated-computer source animacy condition while the right panel reflects the stated-human condition. Each point represents a participant's proportion of words recalled. Diamonds represent the overall mean for each condition.

Cumulative-link regression was used to test if clarity ratings differed as a function of both item- and source-animacy. Animate words were judged as clearer than inanimate words ($\beta = 0.338$, $SE = 0.063$, Wald $Z = 5.363$, $p <$ .001). Despite both source conditions receiving identical stimuli, there was a significant difference in perceived clarity across the two source conditions, such that participants in the stated-computer condition judged the words they heard as clearer than those in the stated-human condition ($\beta = 0.297$, $SE = 0.063$, Wald $Z = 4.714$, $p < .001$). Regarding the manipulation check, the analyses revealed no significant differences between source-animacy conditions in the extent that participants believed the cover story ($\beta = 0$, $SE = 0.2637$, Wald $Z = 0$, $p = .999$). Participants who were told that the items were produced by a computer accepted this story to a similar degree as those who were told the voice was human. The median belief across conditions ($Mdn = 3$) suggests a moderate belief in the manipulation, with perhaps some degree of uncertainty.

Mixed-effects logistic regression models that predicted recall success of each word were used to test if any control variables could account for the observed item-animacy effect. The baseline model included participants as random intercepts and source-animacy, item-animacy, and their interaction as fixed effects. Source-animacy was not allowed to vary as random slopes, as in Experiment 1, because it was not a significant predictor of recall and did not alter the subsequent pattern of results. Control variables were individually entered into the baseline model as a fixed effect. Clarity ratings for each item ($\beta = 0.268$, $SE = 0.046$, Wald $Z = 5.794$, $p < .001$) were a significant predictor of recall success, such that recall was more likely for items with higher clarity ratings. While including clarity ratings into the model did not alter the observed item-animacy effect, the model did account for significantly more variance in recall than the baseline model, $\chi^2(1, N = 1) = 36.615$, $p < .001$. Participants' belief in the cover story ($\beta = 0.063$, $SE = 0.05$, Wald $Z = 1.233$, $p = .217$) was not a significant predictor of recall, did not alter the significance of the item-animacy effect, and did not significantly improved the model's ability to account for variance in recall, $\chi^2(1, N = 1) = 1.506$, $p = .22$.

The present experiment failed to find evidence of a source-animacy effect. It is important to note that participants were not actually exposed to an animate source, and instead those in the stated-human condition were told an inanimate source was animate. This result provides support for the hypothesis that some intrinsic qualities of the auditory signal are necessary for a source-animacy effect to accrue and suggests a boundary condition for the source-animacy effect: beliefs about the animacy of sources alone do not confer mnemonic benefits. This appears congruent with the evolutionary argument that the human voice has a special status in information processing, which may have been selected for by similar evolutionary forces that gave rise to the item-animacy effect. The present experiment provided an additional replication of the item-animacy effect within both an auditory presentation modality and an incidental encoding task.

General Discussion

The key finding of Experiment 1 was that words presented by the human-voiced source were better remembered than words presented by the computer-voiced source. This novel result suggests that the animacy of the source, and not only of the word presented, influences recall. The item-animacy effect was also replicated in an auditory modality. While prior work has explored auditory presentation of nonwords paired with animate or inanimate characteristics (Aslan & John, 2016) or auditory presentation of items followed by a recognition memory test (Stori et al., 2018), the present extension of the item-animacy effect demonstrated that it can also be observed with auditorily presented words and a free recall assessment, which is consistent with the evolutionary explanation of the animacy effect (Bonin et al., 2014; Nairne et al., 2013), as human speech emerged before written communication.

Experiment 1 demonstrated that animacy not only influences the memorability of items, but also the memorability of items presented by an animate source. One possible explanation of this source-animacy effect may be a contagion mechanism (Rozin et al., 1986), where the animacy of the source confers a mnemonic benefit to the information presented by it through association. A second possible explanation is that the human voice holds a special status in memory (Weiss et al., 2012), which may have been conferred through natural selection and may possibly extend to other animate sources. While the present experiments were not intended to disambiguate between these two explanations, future work should attempt to uncover its underlying mechanism.

Though not a direct replication of Experiment 1, Experiment 2 also examined source-animacy using an auditory modality. This experiment explored whether the mnemonic benefit of animate sources is determined by belief about animacy independent of the auditory signal or if it is determined by intrinsic qualities of the auditory signal itself. To this end, participants were presented with a single voice and their belief about whether it was from a human-voiced or computer-voiced source was manipulated. No differences in recall were found under these conditions, which provides support for the intrinsic qualities hypothesis: the human voice may be necessary for the source-animacy effect to emerge and that belief about the source's animacy is not sufficient for the effect to emerge. The necessity of the human voice may arise from either perceptual expertise with human voices or a particular biological significance. Under this hypothesis, the computer-generated voice in Experiment 2 would be treated fundamentally differently than a human voice regardless of what participants are told, or believe, about the source. One possible alternative explanation to this is that the suggestion was not strong enough for participants in the stated-human condition to treat the computer-generated voice in the same way they would a human voice. A stronger suggestion could be provided to increase belief in the manipulation and possibly give rise to a source-animacy effect. With this alternative explanation in mind, future research is warranted to further disambiguate these possible accounts.

Both experiments included additional analyses to mitigate possible alternative explanations. While pleasantness and clarity differed between source-animacy conditions, they were not related to recall performance and could not explain either of the observed item- or source-animacy effects. The divergence in clarity ratings between source-animacy conditions were not related to recall, which is theoretically interesting. There is some research suggesting a desirable difficulty effect in memory such that difficult-to-perceive words are better remembered (Rosner, Davis, & Milliken, 2015). Besken and Mulligan (2014) provided evidence supporting the benefits of desirable difficulties by demonstrating that aurally-distorted words were better remembered on a free recall assessment than non-distorted words. Experiment 1 results showed, however, that although the computer-voiced words were judged as less clear, they were not better recalled, inconsistent with a desirable difficulty effect. It is possible that the source-animacy effect overwhelmed any benefits of perceptual dis-fluency.

The results of Experiment 2 further complicate the role of perceptual clarity. Participants in the stated-computer condition rated the words they heard as significantly clearer than those in the stated-human condition. This is despite that the voices used were identical in both conditions. Though the results of Experiment 2 suggest that while beliefs might play a negligible role in a possible source-animacy effect, they may influence how people judge perceptual clarity. Whatever the case, the results suggest that clarity differences between the voices cannot account for the mnemonic benefit of human-spoken words.

Despite the noteworthy finding of the source-animacy effect and the replication of the item-animacy effect in Experiment 1, it is necessary to consider some important limitations. First, the materials were recorded using a limited number of voices. In order to ensure that these findings are generalizable, future studies must use a wider variety of voices, both computer-generated and human. Second, Experiment 1 rested on an experimenter-defined source-animacy manipulation without testing whether participants viewed the human-voiced source as more animate than the computer-voiced source. Third, while the stimuli were tested for perceptibility by a single research assistant, it is possible that participants may have had more difficulty in perceiving each word. Future work would benefit from more robust norming of the animacy and perceptibility of the auditory sources. Fourth, as Experiment 2 was a between-subject manipulation, participants may have anchored their clarity ratings differently based on whether they were told it was from a computer or human, which may have biased the clarity ratings and obscured a potential relationship between clarity and source-animacy condition. To address this concern, future work should provide a fixed reference for participants to evaluate clarity with respect to. Finally, though clarity and pleasantness were found not to affect the main findings, there may have been other potentially nontrivial differences in vocal variables (e.g. tempo, pitch) that were not recorded or analyzed in the present experiments.

References

Aslan, A., John, T. (2016). The development of adaptive memory: Young children show enhanced retention of animacy-related information. *Journal of Experimental Child Psychology, 152,* 343-350.

Audacity [Computer Software]. (2014). Retrieved from https://www.audacityteam.org/

Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(2),* 429.

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition, 42(3)*, 370-382.

Cogdill, M., Nairne, J., & Pandeirada, J. (2016). Enhanced retention for objects touched by agents. Poster presented at the 28th Annual Convention of the Association for Psychological Science, Chicago, IL.

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. Hirschfeld, S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85-116). Cambridge, England: Cambridge University Press.

Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, UK: Oxford University Press.

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. (2007). Adaptive memory: survival processing enhances retention. Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2),* 263-373.

Nairne, J. S., VanArsdall, J. E., Cogdill, M., (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science, 26(1),* 22-27.

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science, 24(10),* 2099-2105.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods, 162(1),* 8-13.

Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica, 160,* 11-22.

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology, 50(4),* 703-712.

Stori, D., Zaar, J., Cooke, M., & Mattys, S. L. (2018). Sound specificity effects in spoken word recognition: The effect of integrality between words and sounds. *Attention, Perception, & Psychophysics, 80,* 222-241.

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N., & Cogdill, M. (2015). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory, 23(5),* 657-663.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science, 10(2),* 159-164.

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science, 23(10),* 1074-1078.

Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers, 20(1)*, 6-10.