Reviewed Preprint
v1 • June 12, 2026
Not revised

Reviewed Preprint
v2 • August 4, 2026
Revised by authors

Perceiving animacy in 'identical' images

Tal Boger, **Chaz Firestone**
Johns Hopkins University, Baltimore, United States

For correspondence:
tboger1@jhu.edu

Competing interests: No competing interests declared

Reviewing editor: Xilin Zhang, Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, South China Normal University, China

eLife Assessment

This **important** study uses an elegant visual-anagram approach to test whether perceived animacy shapes visual working memory and guides visual attention while tightly controlling for lower- and mid-level image properties. The evidence is **convincing** and provides a rigorous demonstration that perceived animacy influences visual cognition beyond the contribution of its typical visual correlates. The findings will be of broad interest to researchers studying high-level vision, attention, and working memory.

https://doi.org/10.7554/eLife.111179.2.sa4

Abstract

Some objects appear 'animate' (e.g., dogs and elephants) while others do not (e.g., boots and sofas). This distinction pervades human cognition, with an expansive literature reporting striking effects of animacy on vision, memory, social perception, and neural organization. But studies of perceived animacy face a persistent challenge: Objects that differ in animacy tend to differ in many lower-level visual features (e.g., shape, texture, spatial frequency). Thus, it remains controversial whether animacy per se — as opposed to its lower-level correlates — drives visual processing. Here, we achieve previously unattainable levels of experimental control to demonstrate that the visual system represents animacy itself, beyond its lower-level covariates. We vary animacy while holding nearly all lower-level features constant by exploiting "visual anagrams" — a diffusion-based technique for generating static images whose interpretations change radically with orientation. Eight pre-registered experiments leverage this approach to demonstrate that representations of animacy structure visual working memory and guide visual attention. Thus, the visual system extracts animacy itself, beyond its lower-level correlates.

Introduction

Among the most consequential distinctions in human cognition concerns an object's *animacy*: Even when motionless, some objects appear animate (e.g., tigers and bears), while others do not (e.g., rocks and sticks). Representing animacy is of critical importance, helping us recognize threats, find friends or caretakers, predict behavior, and more generally determine which entities in the world have beliefs, desires, and goals. Indeed, researchers across many disciplines have hypothesized that animacy is an ancestrally prioritized feature that may be privileged in the mind [1]. Accordingly, an expansive psychological literature has catalogued numerous effects of perceived animacy on various mental processes, including visual search [2], change detection [1], working memory [3], word learning [4], and even neural organization [5-7]. For example, it is easier to locate an animate object (e.g., a dog) among inanimate objects (e.g., tables, plates, and chairs) than among other animate objects (e.g., birds, fish, and cats), suggesting that basic processes of visual attention are sensitive to the animacy of visual stimuli [1,8].

However, all such research faces a challenge: Objects that differ in animacy also differ in a variety of additional features, including shape, curvature, texture, and other 'low-level' stimulus properties. Thus, it is difficult to attribute experimental effects to animacy itself as opposed to these lower-level covariates. Indeed, previous work suggests that systematic differences in curvature explain away effects of animacy on search [8], and that shape and texture alone may be sufficient to classify animates [9,10]; such regularities even apply to studies of perceived animacy in macaques [11]. Though some approaches show promise in disentangling animacy from one correlate at a time [12], a comprehensive and generalizable solution remains an unrealized goal.

The present work: Perceived animacy in visual anagrams

Here, we pursue a new approach to studying perceived animacy. Recent advances in generative artificial intelligence enable the creation of 'visual anagrams' [13] — images whose interpretations change from one category to another when rotated. We generated a novel stimulus set of anagrams in ways that varied animacy — for example, a dog in one orientation and a boot when rotated, or an elephant in one orientation and a sofa when rotated. Whereas natural images (or line drawings) of dogs and boots differ in animacy, they also differ in shape, curvature, and texture. However, when a visual anagram looks like a dog in one orientation but a boot in another, it is *the very same image* in both cases — composed of the same pixels, just rotated (Figure 1) — such that nearly all lower-level features remain constant (see Materials and Methods for further discussion) while animacy changes categorically. In this way, visual anagrams enable far greater control of high-level perceptual properties than has previously been possible [14]. Eight pre-registered experiments leverage this approach to isolate perceived animacy, demonstrating effects on perception, attention, and working memory.

Results

Experiment 1 used a visual working memory task (Figure 2A). On each trial, participants viewed an array of 5 unique anagrams (e.g., airplane, horse, rabbit, elephant, car). After a 3-second retention interval, one of the images was replaced with a new anagram that was held out of the original 5 (e.g., dog/boot), and participants had to indicate which item had changed by clicking on its position. Crucially, changes either altered the item's animacy (e.g., rabbit→boot) or did not (e.g., rabbit→dog). We observed higher accuracy for changes that altered animacy vs. those that did not (M = 3.7%, $t(27) = 2.36$, $p = 0.025$, $d = 0.45$); in other words, rabbit→boot was more detectable than rabbit→dog, even though (a) the boot and dog are the very same image (just rotated), and (b) animacy itself was completely task-irrelevant (and was never mentioned to subjects).

Importantly, Experiment 1 contained a 'labeling' phase where participants viewed all the anagrams and matched them to their corresponding labels before the start of the task (to ensure the anagrams were recognized as intended). Experiment 2 repeated Experiment 1 without this phase, and found the same pattern: Participants were more accurate for changes that altered animacy than those that did not (M = 4.7%, $t(27) = 2.99$, $p < .01$, $d = 0.57$), internally replicating our key result and ensuring that the labeling phase did not exert undue bias on participants.

Might animacy also affect more canonically *visual* processes, such as attention? This possibility is contested in the literature; whereas some studies find effects of animacy on visual search [2], others suggest that facilitated search is driven merely by differences in curvature between animates and inanimates [8]. In Experiments 3–4 (Figure 2B), we adapted a visual search task from Long et al. (2017). On each trial, participants previewed one anagram image (their target) and then saw an array of 6 unique anagrams; the task was simply to locate the previewed target within this array (and indicate its location by pressing a number key corresponding to its position). Half of trials contained 'uniform animacy' search arrays (where all objects were either animate or inanimate), while the other half contained 'mixed animacy' arrays (where an animate target was surrounded by inanimate distractors, and vice versa). We observed faster search on mixed animacy trials (M = 108ms, $t(28) = 7.04$, $p < .001$, $d = 1.31$; Experiment 3) — even though the mixed animacy and uniform animacy trials contained the exact same images (just rotated). As before, this effect replicated without the labeling phase (M = 124ms, $t(29) = 7.10$, $p < .001$, $d = 1.30$; Experiment 4).

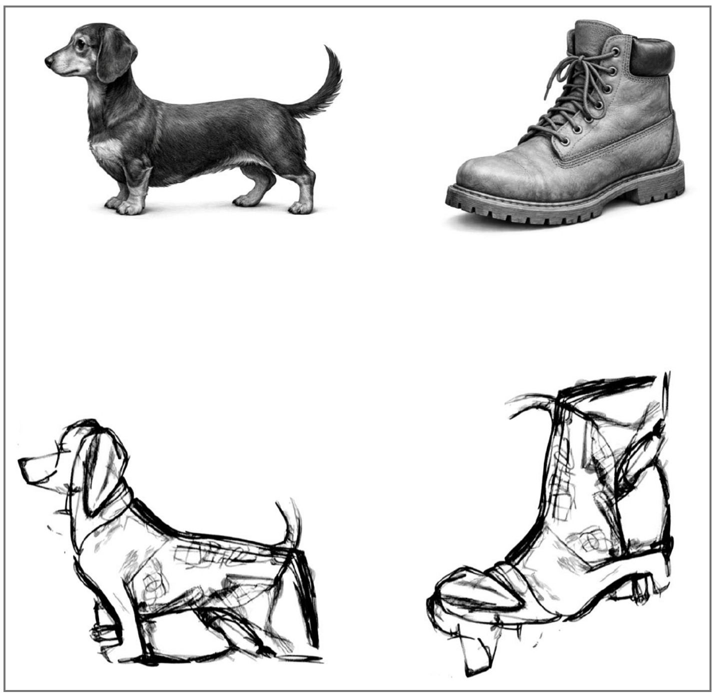

Fig. 1.

(Top Row) This dog and boot differ in animacy, but they also differ in many lower-level features. (Bottom Row) This dog and boot are visual anagrams: They differ in animacy while preserving nearly all lower-level features, because they are the very same image rotated 90°.

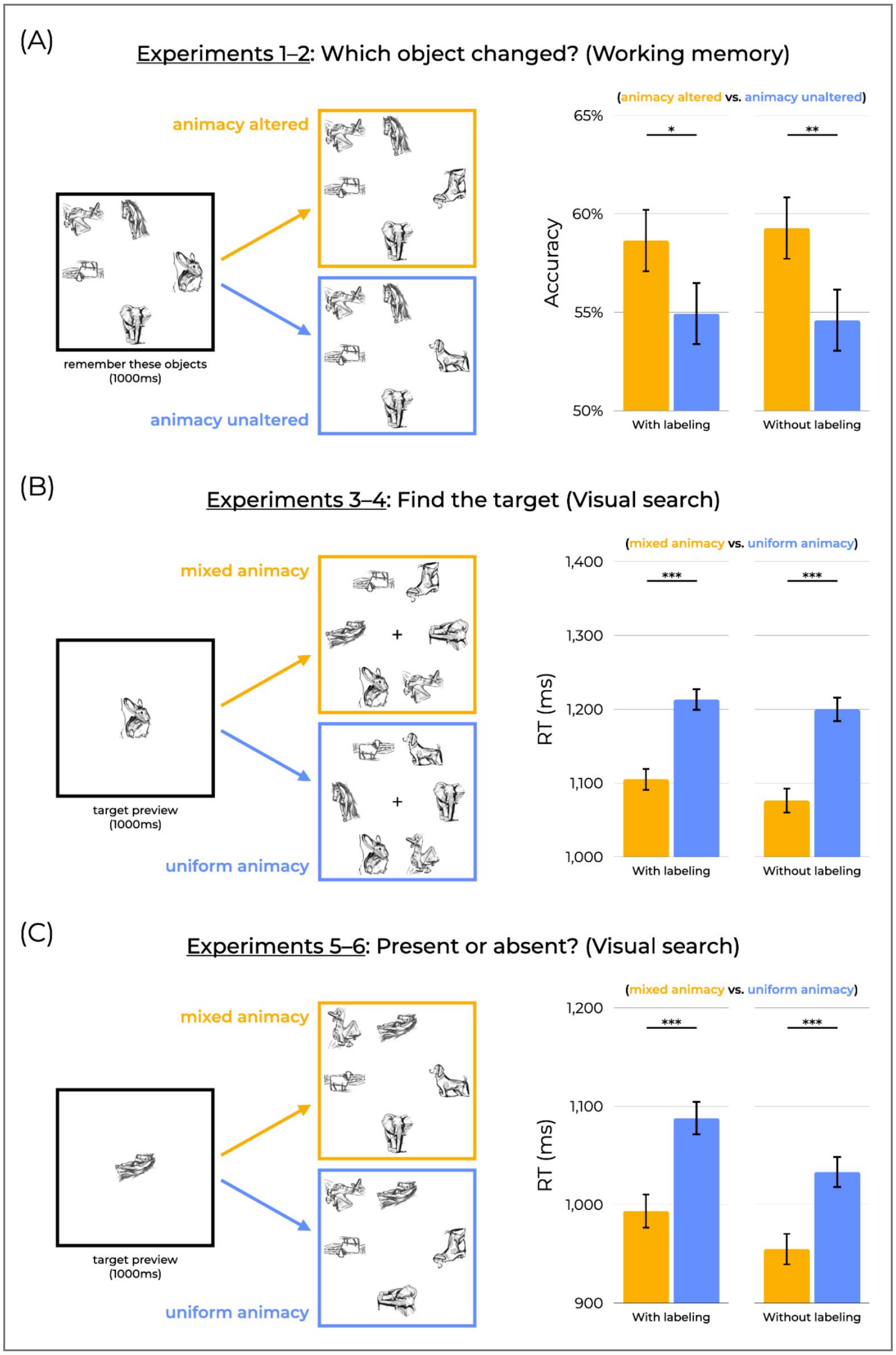

Fig. 2.

(A) In a visual working memory task, changes that altered animacy (e.g., rabbit→boot) were more detectable than changes that did not (e.g., rabbit→dog), even though animacy was task-irrelevant. (B) In a visual search task (modeled after [2]), participants were faster to find targets in mixed animacy arrays than in uniform animacy arrays. (C) This search effect also arose in a present/absent search task. Error bars represent standard errors of the differences across conditions. Readers may view our tasks at *https://perceptionresearch.org/anagrams_animacy*.

Experiments 5–6 (Figure 2C) further tested effects of animacy on basic mechanisms of visual attention by using a more conventional visual search task. Here, participants indicated whether a target object was present or absent in a search array (uniform animacy or mixed animacy). As before, search was faster on mixed animacy trials than uniform animacy trials (M = 94.4ms, $t(29)$ = 5.37, p < .001, d = 0.98; Experiment 5) — an effect that also held without the labeling phase (M = 78.5ms, $t(28)$ = 4.80, p < .001, d = 0.89; Experiment 6).

In each of the above cases, our observed effects cannot be attributed to the lower-level features researchers typically worry about in similar studies (e.g., shape, texture, etc.), because orientation is the only feature that changed across the two interpretations of each anagram. But changes in orientation are of course changes to an image's low-level features; what if that difference explains our results? In other words, perhaps our previously observed effects merely reflect search advantages for orientation, aspect ratio, or center of mass. Though there is little independent reason to suppose that animates and inanimates are, as a group, distinguished by orientation or aspect ratio, we nevertheless addressed this issue in a follow-up experiment. Experiment 7 (Figure 3A) repeated the present/absent search task from Experiments 5–6 but contained a block of trials that depicted silhouetted versions of the anagrams (via convex hull). These silhouettes are just indistinct blobs (and thus not identifiably animate or inanimate) but they share orientation and aspect ratio information with our original stimuli; thus, they reveal whether orientation alone produces similar effects. We found no search advantage for the silhouetted stimuli (M = 2.16ms, $t(144)$ = 0.17, p = 0.86, d = 0.01, BF10 = 0.09). Crucially, Experiment 7 also included anagram trials of the sort appearing in Experiments 5–6. This allowed us to replicate (once again) the key animacy effects with visual anagrams (M = 75.7ms, $t(144)$ = 6.55, p < .001, d = 0.54), and perhaps most importantly ensure that the *difference* in search advantage between the two stimulus classes (anagrams vs. silhouettes) was itself significant (M = 73.5ms, $t(144)$ = 4.24, p < .001, d = 0.35). This last result ensures that even if there were a small but undetected role for orientation, our effects go beyond its contribution.

In a final experiment (Experiment 8; Figure 3B), we asked whether our search effects merely reflect semantic labeling processes. In other words, linguistic labels of the depicted objects access their semantic representations of animacy; would simply replacing the anagram images with their linguistic labels produce the same search effects? Experiment 8 used the same design as in Experiment 7 but with written labels. Participants completed a present/absent search task. One half of trials presented anagram images, and the other half presented labels (i.e., the text "dog" and "boot", etc., appearing in a search array). Once again, we replicated the previous search advantages for anagrams (M = 114ms, $t(29)$ = 5.04, p < 0.001, d = 0.92). However, we observed no search advantage for the words (M = −6.38ms, $t(29)$ = 0.30, p = 0.77, d = 0.05, BF10 = 0.20); and the difference in search advantages between the anagrams and words was itself significant (M = 120.0ms, $t(29)$ = 3.68, p < 0.001, d = 0.67). Given that linguistic labels for the anagrams fail to produce a significant search effect, our results implicate visual processing of animacy rather than merely any semantically activated representation of animacy.

Discussion

Is visual processing sensitive to animacy itself? Given the centrality of animacy to a range of cognitive processes, this question has motivated much research in psychology and neuroscience. However, it has remained without a conclusive answer because animacy is tightly (and seemingly intrinsically) entangled with shape, curvature, texture, and other lower-level features. The present work takes a novel approach to this question, finding that even when two images are pixelwise identical in ways that isolate animacy from its lower-level covariates, they are processed differently by basic processes of perception and cognition.

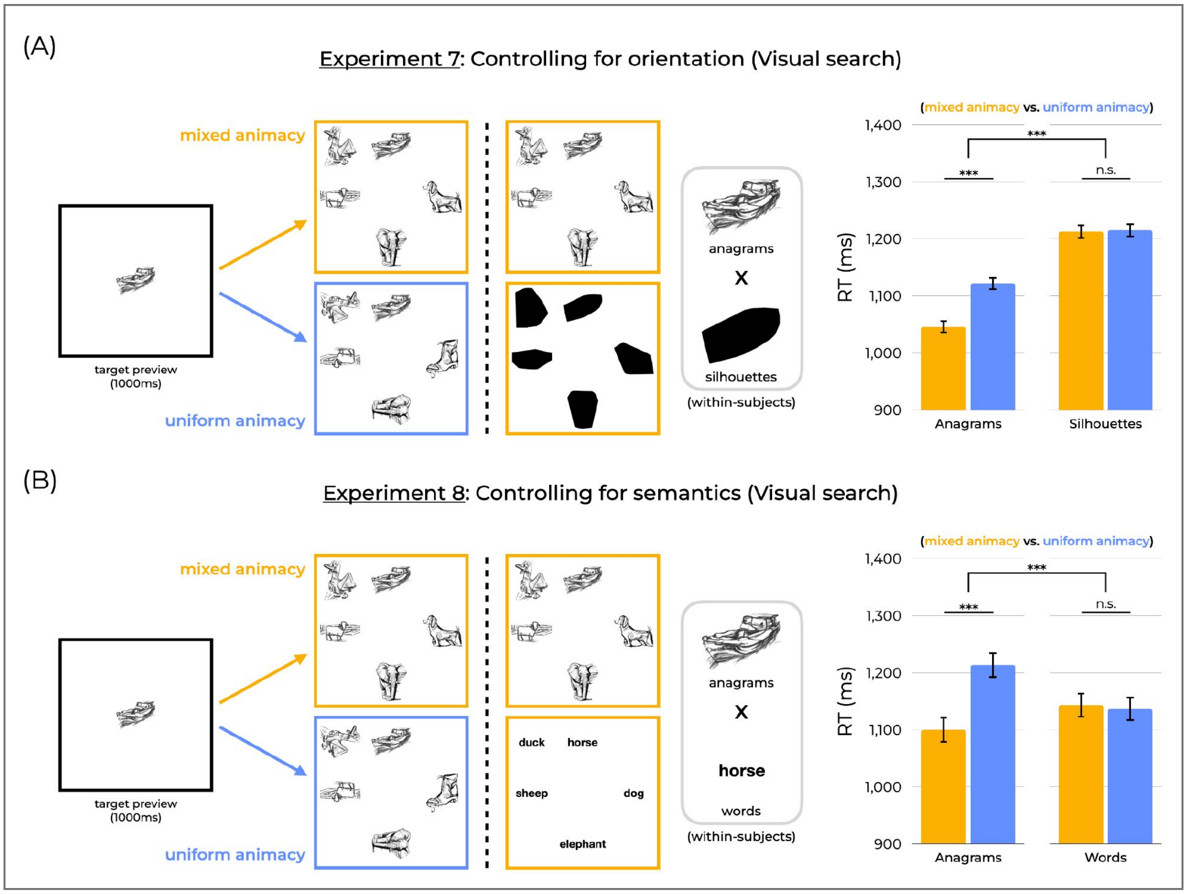

Fig. 3.

(A) To control for orientation (perhaps the only feature left uncontrolled by visual anagrams), we replicated the present/absent search task with silhouetted versions of the anagrams (within-subjects). While anagrams again showed a search advantage, silhouettes did not, suggesting that orientation differences cannot account for our results. (B) To control for linguistic labeling and/or effects of semantics, we replicated the present/absent search task once again with linguistic labels for the anagrams (within-subjects). We observed a significant search advantage for the anagrams, but not for the words, meaning that linguistic labeling cannot account for our results. Error bars represent standard errors of the differences across conditions.

Our approach builds upon related work exploring the relation between lower-level features and high-level properties. For example, one popular approach involves 'texforms', a texture-synthesis algorithm that transforms stimuli into unrecognizable blobs, disrupting high-level recognition but preserving mid- and low-level features [2,6,15]. Our approach provides a complement, isolating animacy from its lower-level correlates rather than the other way around.

Our findings are also consistent with theoretical frameworks emphasizing the importance of animacy in human cognition. Such theories suggest that the centrality of animacy in our lives (both in modern times and ancestrally) has led the mind to prioritize it, producing effects of animacy on basic cognitive tasks including visual change detection [1,16]. Our work offers even stronger and more controlled evidence for many of these conclusions.

Although our approach rules out alternative, lower-level explanations, it remains possible that a high-level property other than animacy — such as natural vs. manmade, which also categorically distinguishes our stimuli — might explain our results. This possibility would still appeal to a (quite related) high-level property and so remain consonant with our general claim. Still, future work could nevertheless address this question directly; for example, rocks, trees, and lakes are natural but inanimate, and so could serve as stimuli in similar experiments.

Finally, the present work interfaces with both longstanding and contemporary philosophical discussions of high-level perception. As Block [17] writes: "What is a high-level representation? Is it just a matter of associations among low-level properties?". Our work suggests that the answer to this latter question is "no". At least for the case of animacy, high-level perception can be disentangled from its lower-level correlates.

Methods

Supplementary methods appear in SI Appendix. Studies were approved by the Johns Hopkins Institutional Review Board; participants gave informed consent.

Materials and methods

General Methods

Open Science Practices

All experiments reported in this paper were pre-registered before data collection. Pre-registrations cover experimental design, sample size, exclusion criteria, and analysis plans. Interested readers may view all experiments at *https://perceptionresearch.org/anagrams_animacy*. An OSF repository containing all data, pre-registrations, analysis scripts, and experiment scripts is available at *https://osf.io/rw8tu/*.

Participants

All participants were recruited via the online platform Prolific (for a discussion of the reliability of this participant pool, see Peer et al., 2017). Each study recruited unique participants.

Stimuli

We used the "visual anagrams" model presented in Geng et al. (2024) to generate our stimuli. (Images were generated via cluster jobs on the Advanced Research Computing at Hopkins core facility.) Stimulus generation proceeded as follows: First, we created a list of animate and inanimate object labels. Then, we generated images reflecting combinations of every pair of objects (drawing one from each list) under three different degrees of rotation (clockwise 90°, counterclockwise 90°, and 180°). This mirrors the stimulus generation approach used in Boger and Firestone (2025), though those stimuli were generated to vary real-world size with orientation (rather than animacy).

Several thousand images are generated as a result and then sorted to select anagrams that were uniquely identifiable as both prompts (before running any experiments). This yielded 12 images appearing in 6 anagram pairs: dog-boot, duck-airplane, elephant-sofa, sheep-car, rabbit-shoes, and horse-boat. Stimuli are available in our OSF repository. This approach controls for nearly all lower-level covariation associated with changes in animacy, allowing for more exacting control than is offered by conventional approaches. It also controls for relevant mid-level features that remain constant with orientation (e.g., configural shape, curvature, etc., which in some contexts are sufficient to produce differences in perceived animacy; e.g., Long et al., 2017; Schmidt et al., 2017; Zachariou et al., 2018; Yetter et al., 2021) — and our specific designs even preserve the relevant ensemble statistics of the displays across trial types. (The anagram pairs are pixelwise-identical subject to rotation, which means that orientation and aspect ratio can change when rotating the anagrams. However, Experiment 7 ensured that such changes cannot account for our observed effects.)

Experiments 1–2: Visual working memory

Experiments 1–2 asked whether animacy affects visual working memory, using a change detection task. Our key question was whether changes that alter an object's animacy (e.g., replacing a rabbit with a boot) are more detectable than changes that do not alter an object's animacy (e.g., replacing a rabbit with a dog — even when the dog and the boot are the same image, just rotated). The pre-registration for Experiment 1 is available at *https://aspredicted.org/uu7r5t.pdf*; the pre-registration for Experiment 2 is available at *https://aspredicted.org/km4nj2.pdf*.

Stimuli and procedure

Participants were instructed to remember an array of objects and then detect any changes to the array. On each trial, the initial array contained 5 unique objects, such that (a) no anagram pairs repeated within an array and (b) one anagram pair could be held out as the 'changing' object. This initial array was visible for 1000ms, after which it disappeared and a 3000ms retention interval began. Following this retention interval, a new array appeared, where one object from the initial array changed identity. The array remained on screen until the participant detected which object changed. Participants responded by first pressing the space bar upon locating the object that changed, and then using the mouse to click on the location of that object.

Participants completed 120 such trials. Each of 6 anagram pairs (i.e., 12 objects) underwent a cross-category (i.e., animacy altered) change 5 times and a within-category (i.e., animacy unaltered) change 5 times. Trial order was randomized for each participant.

The task also contained 8 catch trials which depicted circles of varying colors (instead of the anagrams stimuli). 4 catch trials appeared at the start of the task, and 4 were randomly interspersed throughout.

Experiment 1 was identical to Experiment 2, with one key exception. In Experiment 1, before the task began, participants completed a 'labeling' phase, in which they viewed a gallery of the anagrams used in the task, were shown candidate labels for those images, and had to click on the images corresponding to those labels. This ensured that participants correctly identified each anagram stimulus as its intended object. Experiment 2 did not contain this labeling phase, which (a) allowed us to examine whether the anagrams are readily identifiable without explicit prompting, and (b) ensured that the labeling phase did not bias participants.

Analysis and results

We recruited 30 participants each for Experiments 1–2. In Experiment 1, 2 participants were excluded due to our pre-registered accuracy criterion (which excluded participants who failed to respond correctly on at least 30% of catch trials); and 194/3584 (5.4%) of trials were excluded for having response times which were either too slow (above 5000ms) or too fast (below 300ms).

We observed higher accuracy for cross-category changes than for within-category changes (M = 3.7%, $t(27) = 2.36$, $p = 0.025$, $d = 0.45$), suggesting that visual working memory encodes animacy itself.

In Experiment 2, 2 participants were excluded for accuracy, and 223/3584 (6.2%) of trials were excluded for speed. Here, we replicated the effect observed in Experiment 1: Participants were more accurate on cross-category than within-category changes (M = 4.7%, $t(27) = 2.99$, $p < .01$, d =

0.57).

Experiments 3–4: Visual search

Experiments 3–4 examined whether animacy guides more basic visual processing. The pre-registration for Experiment 3 is available at *https://aspredicted.org/i85iv9.pdf*; the pre-registration for Experiment 4 is available at *https://aspredicted.org/aq2tw5.pdf*.

Stimuli and procedure

Participants completed a visual search task modeled after Long et al. (2017). On each trial, a target object was previewed for 1000ms. The target then disappeared and there was a 500ms inter-stimulus interval (ISI), after which 6 objects appeared, radially positioned around a central fixation cross. The 6 objects each depicted a unique anagram (i.e., none of the objects within a given trial were anagrams of each other). Participants had to search for the previewed target object (which was always among the 6 objects in the search array). Upon finding the target, participants pressed the space bar, after which point all the images were replaced with numbers (arranged 1 through 6, clockwise). Participants used their keyboard to select the number that occupied the location of the target.

Participants completed 144 test trials (in a random order for each participant). Each of 6 anagram pairs (i.e., 12 objects) appeared as the target in each of the 6 locations (1 through 6) twice, once in a “mixed animacy” trial and once in a “uniform animacy” trial. In mixed animacy trials, the target differed in animacy from its distractors (i.e., animate target surrounded by inanimate distractors, or inanimate target surrounded by animate distractors); in uniform animacy trials, the target did not differ in animacy from its distractors (i.e., animate target surrounded by animate distractors, or inanimate target surrounded by inanimate distractors). Additionally, 8 catch trials were included (depicting basic shapes) — 4 at the start of the experiment, and 4 randomly interspersed throughout.

Experiment 3 contained the labeling phase described in Experiment 1, while Experiment 4 did not; otherwise, the two experiments were identical.

Analyses and results

Participants were excluded for failing to respond correctly on at least 75% of catch trials; and trials with response times below 300ms or above 3000ms were excluded. In Experiment 3, 30 participants completed the task, 1 of which was excluded (and 222 of 4408, or 5.0%, of trials were excluded); in Experiment 4, 30 participants completed the task, none of which were excluded (and 165 of 4560, or 3.6%, of trials were excluded).

In both experiments, we observed faster response times for mixed animacy arrays than uniform animacy arrays (Experiment 3: $M = 108$ms, $t(28) = 7.04$, $p < .001$, $d = 1.31$; Experiment 4: $M = 124$ms, $t(29) = 7.10$, $p < .001$, $d = 1.30$). Additionally, this effect did not arise merely because of a speed-accuracy tradeoff, as participants were no less accurate on mixed animacy trials (and if anything were numerically more accurate; Experiment 3: $M = 0.8\%$, $t(28) = 1.34$, $p = 0.19$, $d = 0.25$; Experiment 4: $M = 0.7\%$, $t(29) = 1.05$, $p = 0.30$, $d = 0.19$).

Experiments 5–6: Visual search (present/absent)

While Experiments 3–4 tested effects of animacy on visual search with a task used in previous work (Long et al., 2017), Experiments 5–6 asked whether such effects might arise in a more typical visual search task. Here, we presented participants with a present/absent task. The pre-registration for Experiment 5 is available at *https://aspredicted.org/bn5qj9.pdf*; the pre-registration for Experiment 6 is available at *https://aspredicted.org/5k7ig2.pdf*.

Stimuli and procedure

On each trial, a target object was previewed for 1000ms. Following a 500ms ISI, a search array appeared, presenting 5 objects (randomly positioned in an invisible 3×4 grid). Participants simply had to say whether the target object was present or absent in this array. The task contained 96 test trials (in a randomized order for each participant), equally split between target-present trials and target-absent trials. The 48 target-present trials were themselves equally split, with one half depicting mixed animacy arrays (where the target differs in animacy from its distractors), and the other half depicting uniform animacy arrays (where the target does not differ in animacy from its distractors). Thus, each of 6 anagram pairs (12 objects) was the previewed target in 4 target-absent trials, 2 target-present-mixed-animacy trials, and 2 target-present-uniform-animacy trials.

As before, the task began with 4 catch trials (and 4 additional catch trials were randomly placed throughout the task). Experiment 5 differed from Experiment 6 only in that the former contained the labeling phase described in Experiment 1, while the latter did not.

Analyses and results

Here, we used the same exclusion criteria as in Experiments 3–4: Participants were excluded for responding correctly on less than 75% of catch trials (which excluded 0 participants in Experiment 5 and 1 participant in Experiment 6), and trials with response times below 300ms or above 3000ms were excluded (excluding 38 of 3120 trials in Experiment 5 and 45 of 3016 in Experiment 6).

As in our previous search experiments, search was faster on mixed animacy arrays than uniform animacy arrays (Experiment 5: M = 94.4ms, $t(29) = 5.37$, $p < .001$, $d = 0.98$; Experiment 6: M = 78.5ms, $t(28) = 4.80$, $p < .001$, $d = 0.89$). Similarly, participants were significantly more accurate on mixed animacy arrays in Experiment 5 (M = 3.3%, $t(29) = 2.69$, $p = .01$, $d = 0.49$), and numerically so in Experiment 6 (M = 0.5%, $t(28) = 0.36$, $p = 0.72$, $d = 0.07$).

Experiment 7: Controlling for orientation

Experiment 7 ensured that differences in orientation cannot explain our previously observed effects of animacy on visual attention. In other words, even though the animate and inanimate versions of a visual anagram contain the same pixels, they differ in orientation — and perhaps that difference accounts for our results. This concern is especially acute in the case of visual search, where even very minor visual differences allow targets to stand out from distractors. To address this, we replicated the present/absent search task from Experiments 5–6, but this time included trials that depicted silhouetted versions of the anagrams. If such trials do not show a search advantage (while the anagram trials do), then orientation alone cannot explain our results. The pre-registration for Experiment 7 is available at *https://aspredicted.org/hk9e52.pdf*.

Stimuli and procedure

We generated silhouetted versions of the visual anagrams by converting them into convex hulls with OpenCV. Thus, the silhouetted stimuli retain the aspect ratio and orientation of the original images, but they present as indistinct blobs that are not readily identifiable; in this way, they test for effects of orientation without animacy.

We used the same task timing and design as in Experiments 5–6. However, in this experiment, the 96 test trials were split up differently. One half of trials depicted visual anagrams (as in Experiments 5–6), and the other half depicted silhouetted versions of those anagrams (within-subjects). The two sets of trials were separated into blocks, and block order was randomized for each subject. Each block (of 48 trials each) was then equally split to contain 24 target-absent trials, 12 being target-present-mixed-animacy trials, and 12 target-present-uniform-animacy trials. This design allowed us to (a) replicate our effects from Experiments 5–6, (b) ask whether similar effects arise with silhouettes, and (c) directly compare the two effects in a statistically powerful within-subjects design.

Analyses and results

We recruited 150 participants for this experiment, 149 of which completed the task. Then, we applied the same exclusion criteria as in Experiments 5–6, which excluded 4 participants and 292 of 15080 trials (1.9%).

First, we found that the mixed-animacy search advantage for anagrams replicated again (M = 75.7ms, $t(144) = 6.55$, $p < .001$, $d = 0.54$). Second, we found *no* search advantage for the silhouetted stimuli (M = 2.16ms, $t(144) = 0.17$, $p = 0.86$, $d = 0.01$, BF10 = 0.09). Finally, and perhaps most importantly, the difference in search advantage between the two stimulus classes was significant (M = 73.5ms, $t(144) = 4.24$, $p < .001$, $d = 0.35$). Thus, changes in orientation or aspect ratio alone cannot explain our results.

As above, we examined differences in accuracy in these two conditions. Participants were no less accurate on mixed-animacy anagram trials (and if anything were numerically more accurate; M = 1.8%, $t(144) = 1.88$, $p = 0.06$, $d = 0.16$); and they were also numerically more accurate on mixed-animacy silhouette trials (M = 1.1%, $t(144) = 0.98$, $p = 0.33$, $d = 0.08$).

Experiment 8: Controlling for semantic effects

Experiment 8 asked whether semantic processing alone (independently of perceived animacy) explains our previously observed results. Perhaps the anagram images merely activate a semantic label for animacy, thus explaining our results without necessarily appealing to visual processing of animacy itself. To address this, we used the exact same design as in Experiment 7, but replaced the silhouette trials with words trials (i.e., trials that presented linguistic labels for the anagrams). The pre-registration for Experiment 8 is available at *https://aspredicted.org/73hj6t.pdf*.

Stimuli and procedure

This design was identical to Experiment 7 in every respect, including trial numbers, block structure, etc. The only difference was that instead of containing one block of anagram trials and one block of silhouette trials, this experiment contained one block of anagram trials and one block of words trials. Words described the anagram images (e.g., the picture of a duck was labeled as "duck") in an attempt to roughly match the word length of the two lists.

Analyses and results

We recruited 30 participants for this experiment. We applied the same exclusion criteria as in Experiment 7, which excluded 0 participants and 60 of 3120 trials (1.9%).

As in prior experiments, we observed a search advantage for anagrams (M = 114ms, $t(29) = 5.04$, $p < 0.001$, $d = 0.92$). However, we observed no search advantage for the words (M = −6.38ms, $t(29) = 0.30$, $p = 0.77$, $d = 0.05$, BF10 = 0.20); and the difference in search advantages between the anagrams and words was itself significant (M = 120ms, $t(29) = 3.68$, $p < 0.001$, $d = 0.67$). This suggests that semantic labeling cannot account for our results.

Participants were numerically more accurate on the mixed-animacy anagram trials (M = 0.6%, $t(29) = 0.24$, $p = 0.82$, $d = 0.04$). Participants were also numerically more accurate on the mixed-animacy words trials (M = 1.2%, $t(29) = 0.51$, $p = 0.61$, $d = 0.09$).

Learning effects?

Upon the suggestion of a reviewer, we report exploratory analyses to ask whether learning or familiarity effects could explain our key findings. We analyze this in two ways. First, we ask whether our key effects arise early, by analyzing the first 10 trials of each type in each experiment (e.g., the first 10 mixed animacy vs. uniform animacy trials in the search experiments). Second, we ask whether the effects differ between the first 10 trials of each type and the last 10 trials of each type. We note that our experiments were not designed with the below analyses in mind, which may be underpowered for the questions they seek to answer. However, they do overall suggest that learning is not the explanation of our findings.

We found evidence that our key effects arise early. In 5/6 experiments (Experiments 1, 3, 4, 5, and 6), the effect was significantly (or in one case marginally) present even at the beginning of the experiment (Experiment 1, $p = 0.07$; Experiment 3, $p = 0.01$; Experiment 4, $p < 0.001$; Experiment 5, $p < 0.001$; Experiment 6, $p < 0.01$), and most of these results survive correction for multiple comparisons. In only one experiment (Experiment 2) was there a numerical disadvantage, but it was not significant (p = 0.34). (We refrain from analyzing Experiments 7–8 here because the concern of the learning effects does not apply to silhouettes or words, as those stimuli did not produce a search effect in the first place.)

Second, we found little to no difference between the start of the task and the end of the task. 4/6 experiments revealed numerically weaker effects at the end of the task than the start, though in each case the differences were not significant (Experiment 1, p = 0.63; Experiment 3, p = 0.76; Experiment 4, p = 0.57; Experiment 5, p = 0.08). 2/6 experiments revealed numerically stronger effects at the end of the task than the start, though only one was significant and would not survive a correction for multiple comparisons (Experiment 2, p = 0.01; Experiment 6, p = 0.23). Together, these analyses suggest that learning and familiarity are unlikely to drive our key effects.

Data availability

All anonymized data, analysis scripts, pre-registrations, and stimuli are available at *https://osf.io/rw8tu/*.

Additional information

Author ORCID iDs

Tal Boger: https://orcid.org/0000-0002-5476-8416
Chaz Firestone: https://orcid.org/0000-0002-1247-2422

References

[1] **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 https://doi.org/10.1073/pnas.0703913104 | PubMed

[2] **Long B,** Störmer VS, Alvarez GA (2017) Mid-level perceptual features contain early cues to animacy. *Journal of Vision* **17**:20-20 https://doi.org/10.1167/17.6.20 | PubMed

[3] **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 https://doi.org/10.3758/s13421-013-0368-8 | PubMed

[4] **Laurino J,** Kaczer L (2019) Animacy as a memory enhancer during novel word learning: Evidence from orthographic and semantic memory tasks. *Memory* **27**:820-828 https://doi.org/10.1080/09658211.2019.1572195 | PubMed

[5] **Konkle T,** Caramazza A (2013) Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience* **33**:10235-10242 https://doi.org/10.1523/jneurosci.0983-13.2013 | PubMed

[6] **Long B,** Yu CP, Konkle T (2018) Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences* **115**:E9015-E9024 https://doi.org/10.1073/pnas.1719616115 | PubMed

[7] **Thorat S,** Proklova D, Peelen MV (2019) The nature of the animacy organization in human ventral temporal cortex. *eLife* **8**:e47142 https://doi.org/10.7554/eLife.47142 | PubMed

[8] **Levin DT,** Takarae Y, Miner AG, Keil F (2001) Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics* **63**:676-697 https://doi.org/10.3758/bf03194429 | PubMed

[9] **Schmidt F,** Hegele M, Fleming RW (2017) Perceiving animacy from shape. *Journal of Vision* **17**:10-10 https://doi.org/10.1167/17.11.10 | PubMed

[10] **Zachariou V,** Del Giacco AC, Ungerleider LG, Yue X (2018) Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *Journal of Vision* **18**:3-3 https://doi.org/10.1167/18.12.3 | PubMed

[11] **Yetter M,** Robert S, Mammarella G, Richmond B, Eldridge MA, Ungerleider LG, Yue X (2021) Curvilinear features are important for animate/inanimate categorization in macaques. *Journal of Vision* **21**:3-3 https://doi.org/10.1167/jov.21.4.3 | PubMed

[12] **Proklova D,** Kaiser D, Peelen MV (2016) Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of Cognitive Neuroscience* **28**:680-692 https://doi.org/10.1162/jocn_a_00924 | PubMed

[13] **Geng D,** Park I, Owens A (2024) Visual anagrams: Generating multi-view optical illusions with diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 24154-24163 https://doi.org/10.1109/cvpr52733.2024.02280

[14] **Boger T,** Firestone C (2025) Visual anagrams reveal high-level effects with 'identical' stimuli. *Current Biology* **35**:R910-R912 https://doi.org/10.1016/j.cub.2025.08.036 | PubMed

[15] **Freeman J,** Simoncelli EP (2011) Metamers of the ventral stream. *Nature Neuroscience* **14**:1195-1201 https://doi.org/10.1038/nn.2889 | PubMed

[16] **Altman MN,** Khislavsky AL, Coverdale ME, Gilger JW (2016) Adaptive attention: How preference for animacy impacts change detection. *Evolution and Human Behavior* **37**:303-314 https://doi.org/10.1016/j.evolhumbehav.2016.01.006

[17] **Block N** (2023) *The border between seeing and thinking* Oxford University Press.

Peer reviews

Reviewer #1 (Public review):

[Editors' note: this version has been assessed by the Reviewing Editor without further input from the original reviewers. The authors have addressed the comments raised in the previous round of review.]

This is a very cool paper that casts light on a persistent problem in the psychology and philosophy of visual representation: is there high-level perception? Every vision scientist agrees that low-level features such as shape, color, texture, motion and spatial frequency are represented in visual perception, but there is a great deal of controversy about the representation of high-level properties such as causation, faces, agency and animacy. Animacy is especially problematic because there are large differences in line curvature between stimuli that represent animate and inanimate items.

This article uses a novel approach-visual "anagrams" that are exactly the same image, except one is rotated 90 degrees relative to the other. They found persistent differences in visual processing between animate and inanimate stimuli. (Of course, the stimuli aren't animate-they represent animate items.). For example, there were processing differences between changes between animate and inanimate items (rabbit to boot) that were not present in rabbit to dog. They also showed such differences in two kinds of visual search tasks.

Of course, there are feature differences that exploit orientation. A classic example is the difference between a square and a diamond that is produced from the square by rotating it 45 degrees.

They addressed an aspect of this challenge having to do with some features using silhouettes. There was no search advantage for silhouetted stimuli.

https://doi.org/10.7554/eLife.111179.2.sa3

Reviewer #2 (Public review):

Summary:

The authors present a creative approach using visual anagrams matched on low-level image statistics to isolate animacy from low-level visual features and report consistent effects of animacy on visual working memory and attention.

Strengths:

(1) An important methodological advance in controlling low-level confounds that have historically complicated the study of animacy.

(2) The converging effects across multiple experiments, together with the pre-registered design, strengthen the reliability of the reported findings.

https://doi.org/10.7554/eLife.111179.2.sa2

Reviewer #3 (Public review):

This study makes clever use of generative AI to create stimuli that are pixel-for-pixel identical but which have radically different meanings depending on their orientation, to investigate the perception of animacy while retaining control over low-level image features (so-called 'anagram' stimuli).

The authors present seven elegantly designed experiments in a commendably compact format.

Experiments 1 and 2 involved a working memory paradigm in which participants had to spot which of five objects in an array changed after a pause. Importantly, the changed object was an anagram stimulus that in one orientation matched the animacy/inanimacy of the changed object, and in the other orientation was the opposite (e.g., a rabbit is replaced by either a dog or a boot, where the dog and boot stimuli are actually identical, just rotated by 90 degrees). They found a difference in accuracy depending on whether the animacy of the objects matched.

Experiments 3 and 4 used a visual search task in which the participants had to localize the target, and the distractors were anagrams that either matched the target in terms of animacy or did not. There was a significant cost in terms of response time when the animacy of the target was the same as that of the distractors. Experiments 5 and 6 also used a similar visual search design, except that the task was to determine if the target was present or absent from the display, and the distractors again either matched or differed from the target in terms of animacy. Again, the authors found slower responses when the distractor arrays matched the animacy of the target than when they differed.

An obvious potential concern about the studies is addressed by Experiment 7. It is unclear if the observed effects are related to the specific orientations of the target and distractor stimuli selected in each condition. For example, it could be that all the animate versions of the anagrams involved tall and skinny shapes, while all the inanimate versions involved wide and short objects, due to the 90-degree rotational difference between the two versions of the stimuli. To control for this, the authors repeated the visual search experiment but with convex-hull silhouettes of each of the stimuli. In other words, all targets and distractors from each trial were replaced by a black splotch with approximately the same overall outline (envelope) as the corresponding stimulus. Importantly, in contrast to the anagram stimuli, the silhouettes had had no meaningful semantic interpretation, and their animacy did not change depending on their orientation.

https://doi.org/10.7554/eLife.111179.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***eLife Assessment***
>
> *This valuable study uses an elegant visual-anagram approach to test whether perceived animacy structures visual working memory and attention while controlling for many low-level image properties. The evidence is solid, with converging results across seven preregistered experiments, but the central claim that animacy itself is represented independently of visual features should be tempered, as residual mid-level configural cues, ensemble or category structure, and broader semantic differences may also contribute to the effects. The work will be of interest to researchers studying high-level visual representation, attention, and working memory.*

We thank the Editors and Reviewers for this careful and informed assessment. We appreciate that every Reviewer found our approach to be elegant, our findings to be solid, and our question to be of broad interest. We respond to each Reviewer's specific comments in more detail below; but we thought to summarize some of the highlights - especially the specific comments that come up in this Assessment - here.

(1) The Reviewers make the insightful point that, even if our stimuli effectively control for many low-level features, there may be other high-level features that explain performance in our experiments (R2: "Although the anagram paradigm effectively controls low-level visual features [...] these stimuli differ not only in animacy but also along other semantic dimensions such as natural versus manmade categories."). We are happy to embrace this possibility. If our results were explained by high-level visual representation of the natural vs. manmade distinction, rather than the animate vs. inanimate distinction, this would still be an appeal to a (not altogether unrelated) high-level property being represented independently from its lower-level features, which was the primary motivation for our study. We framed our work specifically around animacy given the persistent debates regarding perceived animacy, as well as the fact that our stimuli do quite saliently vary along that dimension; but we are certainly open to other nearby high-level categories being at play. We also think this is an empirical question that could be tested in future work. For example, objects like rocks and lakes are natural but inanimate. If they behave more like dogs than like boots in our paradigms, then Reviewer #2 may be right that naturalness was the relevant property all along; but if they behave more like boots than like dogs, then perhaps it really was animacy doing the work. We now discuss this explicitly in our paper, and we appreciate the opportunity to not only clarify our claims but also spur discussion for future work.

(2) Multiple Reviewers raise the question of whether semantic factors that go beyond the images themselves may be driving our effects. Reviewer #3 raises a particularly interesting question along these lines: "if all the stimuli in the experiments were replaced with the verbal names of the depicted objects instead of pictures, would we expect different results?" We have now taken this question quite literally and run this experiment exactly as described. Of course, much research already explores cognitive processing of animate/inanimate words, finding (for example) stronger memory for animate objects than inanimate ones (e.g., Nairne et al., 2013; Nairne et al., 2017). However, such tasks do not invoke effects of visual processing, whereas the question at issue here is specifically whether the visual system prioritizes animacy independent of its lower-level features. To this end, we conducted a new, pre-registered experiment (now Experiment 8) where participants search for animate/inanimate words on some trials, and animate/inanimate pictures on others. Given the nature of visual search tasks, we should expect to find no search advantage for words (as their meanings are not processed in vision per se) — and we should also expect to replicate (once again) our search advantage for pictures. This is exactly what we found. In other

words, linguistic stimuli alone failed to produce the effect, while anagrams did produce the effect. We believe this rules out the strongest form of the semantic labeling account.

(3) Finally, Reviewers #2 and #4 raise some concerns regarding residual mid-level features such as configural shape and ensemble statistics, which lie somewhere between animacy itself and more basic properties like contrast or spatial frequency. In our paper, we now clarify each of these concerns in greater detail. In short: We think that our stimuli and experiments indeed control for these residual cues. For example, rotating an image preserves its configural shape; and, as we argue below, the specific ensemble statistics argument fails to get off the ground without appeal to animacy itself.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *Evidence for visual representation of animacy.*
>
> *Strengths:*
>
> *This is a very cool paper that casts light on a persistent problem in the psychology and philosophy of visual representation: is there high-level perception? Every vision scientist agrees that low-level features such as shape, color, texture, motion and spatial frequency are represented in visual perception, but there is a great deal of controversy about the representation of high-level properties such as causation, faces, agency and animacy. Animacy is especially problematic because there are large differences in line curvature between stimuli that represent animate and inanimate items.*
>
> *This article uses a novel approach-visual "anagrams" that are exactly the same image, except one is rotated 90 degrees relative to the other. They found persistent differences in visual processing between animate and inanimate stimuli. (Of course, the stimuli aren't animate-they represent animate items). For example, there were processing differences between changes between animate and inanimate items (rabbit to boot) that were not present in rabbit to dog. They also showed such differences in two kinds of visual search tasks.*
>
> *Of course, there are feature differences that exploit orientation. A classic example is the difference between a square and a diamond that is produced from the square by rotating it 45 degrees.*
>
> *They addressed an aspect of this challenge having to do with some features using silhouettes. There was no search advantage for silhouetted stimuli.*
>
> *Weaknesses:*
>
> *I thought this was an excellent submission. I have two suggestions for revision:*

We are glad to hear this Reviewer recognizes the broad challenge we are tackling in this work (separating high-level from low-level visual features) and found our submission to be “excellent”.

> *(1) I thought that experiment 7 should have been described in more detail, with the upshot explained better. What exactly do the authors take it to show?*

Sorry for the lack of clarity here. We think the Reviewer actually gets this right earlier in their review; many feature differences exploit orientation, and our silhouettes control (Experiment 7) shows that those differences alone fail to explain our effects. For example, one might worry that our search effects merely reflect oddities in the aspect ratio or center of mass of the images. Converting the anagrams into silhouettes preserves these features. Thus, the fact that we found no search advantage with silhouettes suggests that these features on their own fail to produce the relevant effects; put the other way around, the effects we observed earlier must go beyond those features. We now discuss this in greater depth in our paper.

> *(2) There should be a candid discussion of what the loose ends are and how they might be addressed. It would be good to have some examples like the square/diamond case with some indication of what would address such challenges.*

We agree with this, though we are somewhat limited by the space constraints of the Short Report format. A primary loose end we see is the possibility that high-level properties other than animacy explain our results (as raised by other Reviewers). We have added some discussion of this possibility to the paper.

We would like to thank this Reviewer for their thoughtful feedback.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *The authors present a creative approach using visual anagrams matched on low-level image statistics to isolate animacy from low-level visual features and report consistent effects of animacy on visual working memory and attention. While this is a thoughtful design and is well executed across seven pre-registered experiments, it remains unclear whether the reported effect is truly driven by animacy, as opposed to broader differences in ensemble statistics or semantic structure across the "mixed animacy" versus "uniform animacy" conditions. As such, the interpretation of a "pure" animacy effect may be overstated.*
>
> *Strengths:*
>
> *(1) An important methodological advance in controlling low-level confounds that have historically complicated the study of animacy.*
>
> *(2) The converging effects across multiple experiments, together with the pre-registered design, strengthen the reliability of the reported findings.*

We are glad to hear this Reviewer found our work to be “creative” and believes it offers an “important methodological advance”.

> *Weaknesses:*
>
> *(1) Specificity of the animacy effect vs. category-level ensemble structure*
>
> *The central claim is that animacy itself drives the observed effects. However, the key manipulation ("mixed animacy" versus "uniform animacy") also introduces differences in category-level ensemble structure. For example, in Experiments 1-2, cross-category change detection (e.g., dog to chair) may be easier not because of animacy per se, but because of a change in overall ensemble statistics (Brady & Alvarez, 2011, 2015). In addition, since each display contains five objects (two in one category and three in the other category), cross-category changes may also alter category balance in a way that further facilitates detection. In contrast, within-category changes preserve both ensemble structure and category composition, making them more difficult to detect.*
>
> *Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. Psychological Science.*

> *Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. Journal of Vision.*

We appreciate the opportunity to clarify our claims and the support for them. Our claim is indeed that animacy (or a closely related high-level property; see below) drives our effects, over and above its lower-level correlates — i.e., that the explanation for differences in change detection or search across conditions will invoke a high-level property of the images. As we understand the Reviewer's concern(s), they either (a) are already addressed by our novel methodology, or (b) would still fall perfectly in line with our claim as stated above.

Consider the Reviewer's concern that cross-category change detection "may be easier not because of animacy per se, but because of a change in overall ensemble statistics". Which ensemble statistics change across categories in our stimulus set? Take as an example the case depicted in our figure, where a rabbit changes into either a dog (within-category) or a boot (cross-category). The dog and the boot are the very same image, just rotated; thus, they have the same luminance, curvature, area, spatial frequency, and so on. So if the change from rabbit to dog changes the array's ensemble statistics with respect to any of those properties, it does so in the very same way as the change from rabbit to boot — and yet detection is still better for rabbit → boot than for rabbit → dog. Indeed, for nearly any ensemble statistic, the difference between the rabbit-display and the dog-display will be identical to the difference between the rabbit-display and the boot-display. To engage with the specific cases discussed in the two cited papers (Brady & Alvarez, 2011, 2015): The dog and the boot are the same size (because they are the same image), so average size is identical (just as average luminance, curvature, area, and spatial frequency are identical). And the very few properties left over (e.g., aspect-ratio) are addressed by later experiments.

To be clear: We are not saying that there are no differences in ensemble statistics between the rabbit-display and the dog-display; across those displays, we replace one image with a different image, so there are likely all kinds of corresponding differences in ensemble statistics. The key question is whether that change in ensemble statistics differs across trial types in ways that might explain our effect - i.e., whether there is any difference between the rabbit-display and dog-display that is not also present between the rabbit-display and the boot-display. We don't see how the answer could be yes, at least with respect to the statistics typically considered. A similar logic applies to the search tasks, with the silhouette control (Experiment 7) providing especially strong evidence that certain ensemble statistics or lower-level features cannot explain our effect.

Now, it's possible the Reviewer is referring to properties other than the low-/mid-level properties we mention above. Perhaps, for example, many animate stimuli on a display at one time have a striking collective appearance (all these animals are looking at me!) that lots of inanimate stimuli do not (this might be related to the Reviewer's concern about "category balance"). But as we see it, this explanation just invokes animacy all over again, and so is the sort of explanation we would embrace.

We now say more about this concern in the paper to be as clear as possible about our claims.

> *(2) Limited stimulus set and potential learning effects*
>
> *The relatively small stimulus set (six anagram pairs) and repeated exposure raise the possibility of learning or familiarity effects. Does performance change over time? e.g., are there meaningful differences between early and late trials (e.g., first 10% vs. last 10%)? If such differences are present, they could suggest the development of task-specific strategies or increased efficiency with repeated exposure, rather than stable effects driven by the experimental manipulation itself.*

This is an interesting question, and we recognize this analysis absent from our initial submission. To be fair, stimulus sets of this size are not unusual in change-detection and search tasks, which often involve red, green, and blue squares repeated over the course of several hundred trials. Still, we certainly take the Reviewer's point here and also embrace their analytical approach to addressing it. We've now run the "familiarity effects" analyses the Reviewer suggests (as well as some they did not suggest). The top-level headline is that learning or familiarity effects cannot explain our results, and if anything most of these analyses not only fail to support this alternative account but actively point against it. Below are more details.

First, we worry that the Reviewer's concern about "the development of task-specific strategies … rather than stable effects driven by the experimental manipulation itself" isn't actually addressed by the suggested analysis of comparing the last 10% of trials to the first 10%. One reason for this is simply that it's possible that both mechanisms are at play - i.e., that there is a baseline difference even without any familiarity that is then enhanced by some learning mechanism. (There are other issues as well: For example, one might imagine that participants get quite good at the task during the middle 80% of trials, but then get fatigued at the end. If this were true, then comparing the first 10% to the last 10% of trials could make it seem like there is no learning or familiarity, even if there were such effects. And on top of all this there is just the issue of statistical power, since far fewer trials go into these analyses than into our primary, pre-registered analyses). Nevertheless, we ran the Reviewer's proposed analyses (using the first and last 10 trials of each type, which offers the best chance to find the pattern the Reviewer is concerned about). If anything, this analysis points in the opposite direction to the Reviewer's prediction: 4/6 experiments (Experiments 1, 3, 4, and 5) revealed numerically weaker effects at the end of the task than the start, while only 2/6 experiments (Experiments 2 and 6) revealed numerically stronger effects at the end of the task than the start. Moreover, most of these results were non-significant, with only one marginal result (Experiment 5, p< = 0.08) and one significant result (Experiment 2, $p = 0.01$), and this is before any correction for multiple comparisons, which would make all of these results non-significant. So even though our account could easily accommodate learning effects, it's not clear that they even exist here in any consistent or reliable way.

Second, however, we think a more informative way to answer the Reviewer's question is to ask not about learning over the course of the experiment but rather whether the key effects arise very early in the task. If they do, then any learning effects arising later couldn't fully account for our results. Now, again, these tests are underpowered and only exploratory (to do this analysis properly, we would want to run entirely new experiments designed for this purpose), but we in fact did find evidence that our key effects arise early. In 5/6 experiments (Experiments 1, 3, 4, 5, and 6), the key effect was significantly (or in one case marginally) present even at the beginning of the experiment (Experiment 1, $p = 0.07$; Experiment 3, $p = 0.01$; Experiment 4, $p < 0.001$; Experiment 5, $p < 0.001$; Experiment 6, $p < 0.01$), and most of these results would survive correction for multiple comparisons. (In only one experiment, Experiment 2, was there a numerical disadvantage, but it was not significant; $p = 0.34$.) So even though our experiments were not designed or powered for this purpose, they do seem to suggest that the effects arise even without much familiarity at all.

All told, we think these analyses suggest quite strongly that learning alone fails to explain our key effects. There is no evidence that the effects in general are stronger at the end of the experiment than the beginning (if anything it is the opposite); and there is evidence that most of the effects we investigated can be detected even very early in the experimental sessions. We have added discussion of these new analyses to our manuscript.

> *(3) Role of semantics*

> *Although the anagram paradigm effectively controls low-level visual features, it still relies on high-level semantics (e.g., "dog" vs. "boot"). These stimuli differ not only in animacy but also along other semantic dimensions such as natural versus manmade categories. From a semantic standpoint, it remains unclear whether the observed effects can be uniquely attributed to animacy or whether they reflect broader conceptual distinctions.*

We agree with the Reviewer here. While we feel comfortable interpreting our effects in terms of a high-level property like animacy as opposed to a lower-level property like curvature, it remains possible that the observed effects reflect some other, closely related high-level distinction (like natural vs. manmade). Our primary concern was to tease apart high-level properties from low-level features, which the Reviewer's question does not threaten — if attention and memory are sensitive to the natural/artificial distinction, that's interesting too, and a near neighbor of our actual claim. Still, we agree that this could be addressed, and we even see it as an empirical question testable in future work. Perhaps the most relevant departures between animate/inanimate and natural/manmade include objects like clouds, plants, and rocks — objects that are natural but not "animate" in the sense often used in this literature. If something like our paradigm revealed that rocks behave more like dogs than like boots, that would suggest that naturalness, rather than animacy, was driving the effects; but if rocks behave more like boots than like dogs, that would point to animacy even more strongly. We remain open-minded about this possibility, but it would of course require multiple new experiments with a brand new stimulus set and so goes beyond the present contribution. In any case, we have added a discussion of this issue to the paper and have adjusted our claims accordingly.

> ***Reviewer #3 (Public review):***
>
> *Summary:*
>
> *This study makes clever use of generative AI to create stimuli that are pixel-for-pixel identical but which have radically different meanings depending on their orientation, to investigate the perception of animacy while retaining control over low-level image features (so-called 'anagram' stimuli).*
>
> *The authors present seven elegantly designed experiments in a commendably compact format.*
>
> *Experiments 1 and 2 involved a working memory paradigm in which participants had to spot which of five objects in an array changed after a pause. Importantly, the changed object was an anagram stimulus that in one orientation matched the animacy/inanimacy of the changed object, and in the other orientation was the opposite (e.g., a rabbit is replaced by either a dog or a boot, where the dog and boot stimuli are actually identical, just rotated by 90 degrees). They found a difference in accuracy depending on whether the animacy of the objects matched.*
>
> *Experiments 3 and 4 used a visual search task in which the participants had to localize the target, and the distractors were anagrams that either matched the target in terms of animacy or did not. There was a significant cost in terms of response time when the animacy of the target was the same as that of the distractors. Experiments 5 and 6 also used a similar visual search design, except that the task was to determine if the target was present or absent from the display, and the distractors again either matched or differed from the target in terms of animacy. Again, the authors found slower responses when the distractor arrays matched the animacy of the target than when they differed.*
>
> *An obvious potential concern about the studies is addressed by Experiment 7. It is unclear if the observed effects are related to the specific orientations of the target and*

> *distractor stimuli selected in each condition. For example, it could be that all the animate versions of the anagrams involved tall and skinny shapes, while all the inanimate versions involved wide and short objects, due to the 90-degree rotational difference between the two versions of the stimuli. To control for this, the authors repeated the visual search experiment but with convex-hull silhouettes of each of the stimuli. In other words, all targets and distractors from each trial were replaced by a black splotch with approximately the same overall outline (envelope) as the corresponding stimulus. Importantly, in contrast to the anagram stimuli, the silhouettes had had no meaningful semantic interpretation, and their animacy did not change depending on their orientation.*
>
> *Strengths:*
>
> *The main strength is the elegant use of stimuli that control almost perfectly for low-level image features.*

Thank you for this kind feedback. This summary perfectly captures both our empirical contribution and the claims we are making.

> *Weaknesses:*
>
> *My only real concern about the study is whether the findings truly provide evidence for a high-level visual representation of animacy independent of the low-level stimulus characteristics, or whether, instead, the effects are essentially semantic priming, which is independent of visual processing per se. For example, if all the stimuli in the experiments were replaced with the verbal names of the depicted objects instead of pictures, would we expect different results? Words can also access semantic representations of the animacy of objects, and also don't suffer from low-level visual confounds. It would be helpful to add a discussion of this possibility to the article.*

Wow, we love this question! And so we've now conducted exactly the experiment the Reviewer suggests here. In a new pre-registered study (Experiment 8), we presented participants with a present/absent search task (as in Experiments 5–7). One half of trials consisted of the anagram stimuli (such that we could, once again, replicate the mixed-animacy search advantage); but the other half of trials consisted of the words describing the anagrams (e.g., "dog", "boot", "sheep", "car", etc.). The experiment worked beautifully: We found no effect with the words, but replicated the search advantage with the pictures — and also found a significant difference between the effects elicited by the two stimulus types.

We agree with the Reviewer that this now rules out the possibility that semantic representations alone explain these visual effects. Thank you!

> ***Reviewer #4 (Public review):***
>
> *In this article, the authors investigate whether perceived animacy influences visual processing independently of lower-level visual features by using "visual anagrams." Across seven experiments, they test whether animacy, isolated from many lower-level visual properties, structures visual working memory and guides visual attention. The central claim is that the visual system may represent animacy itself, rather than animacy emerging solely from associations among low-level visual properties.*
>
> *I find this investigation compelling. The experiments described provide strong control over several lower-level visual features, including curvature, texture, and related image properties. However, the visual anagrams are not pixelwise-identical across orientations. Because the images are rotated, the retinal configuration of pixels and the spatial organization of some low- to mid-level shape features also change. As a result, the*

> *configural arrangement of mid-level visual features may still contribute to perceived animacy.*

We are glad to hear the Reviewer finds our investigation “compelling”.

> *I encourage the authors to discuss how independent perceived animacy is in this context from the contribution of mid-level visual features, such as configural shape cues that are diagnostic of animacy. This distinction would help sharpen the interpretation of the results and more precisely define the level of visual representation isolated by the visual-anagram approach.*

This is a helpful point, and it also echoes a sentiment expressed by Reviewer #2. While configural shape is diagnostic of animacy writ large, it can’t account for our observed effects here because rotating an image does not vary its configural shape. We now mention this in our work, and we agree that it helps sharpen the interpretation of our studies.

> *Additionally, previous studies have argued that low- and mid-level curvilinear features may contribute to animate/inanimate categorization, and may in some cases be sufficient to support such distinctions (e.g., PMID: 33798259; PMID: 28654965). I encourage the authors to clarify how these previous findings on curvilinearity and rectilinearity fit with the overarching claim of the current study, namely that the visual system may represent animacy itself rather than animacy emerging solely from associations among lower-level visual properties.*

Yes, many studies from exactly that corner of the field actually motivated the present work, which is why we cited them in our submission. In a way, we are approaching this issue from the other side of the equation. Whereas the papers the Reviewer points to (along with many others) ask whether mid-level features (such as curvilinearity and rectilinearity) are sufficient to support perceived animacy, we ask whether these and other features are necessary to support perceived animacy. Prior work is relatively split on this issue, leaving the question wide open. We take our work to show that differences in curvature are not necessary for differences in perceived animacy, because our anagrams have identical curvature yet differ in animacy — and the visual system capitalizes on that difference. Put the other way around, representation of animacy can and does go beyond representation of its low- and mid-level correlates. Thank you!

https://doi.org/10.7554/eLife.111179.2.sa0