# Mid-level perceptual features contain early cues to animacy

**Bria Long**
Department of Psychology, Harvard University, Cambridge, MA, USA

**Viola S. Störmer**
Department of Psychology, University of California, San Diego, CA, USA

**George A. Alvarez**
Department of Psychology, Harvard University, Cambridge, MA, USA

**While substantial work has focused on how the visual system achieves basic-level recognition, less work has asked about how it supports large-scale distinctions between objects, such as animacy and real-world size. Previous work has shown that these dimensions are reflected in our neural object representations (Konkle & Caramazza, 2013), and that objects of different real-world sizes have different mid-level perceptual features (Long, Konkle, Cohen, & Alvarez, 2016). Here, we test the hypothesis that animates and manmade objects also differ in mid-level perceptual features. To do so, we generated synthetic images of animals and objects that preserve some texture and form information ("texforms"), but are not identifiable at the basic level. We used visual search efficiency as an index of perceptual similarity, as search is slower when targets are perceptually similar to distractors. Across three experiments, we find that observers can find animals faster among objects than among other animals, and vice versa, and that these results hold when stimuli are reduced to unrecognizable texforms. Electrophysiological evidence revealed that this mixed-animacy search advantage emerges during early stages of target individuation, and not during later stages associated with semantic processing. Lastly, we find that perceived curvature explains part of the mixed-animacy search advantage and that observers use perceived curvature to classify texforms as animate/inanimate. Taken together, these findings suggest that mid-level perceptual features, including curvature, contain cues to whether an object may be animate versus manmade. We propose that the visual system capitalizes on these early cues to facilitate object detection, recognition, and classification.**

## Introduction

Most research on object recognition has focused on how the visual system accomplishes basic-level recognition, asking how it is that we see a dog as a dog or a cup as a cup (DiCarlo & Cox, 2007; Grill-Spector & Kanwisher, 2005; Rosch et al., 1976). Indeed, understanding the computations that lead to these basic-level representations is no small task. However, relatively little work has investigated the principles the visual system uses to organize these diverse object representations into broader categories.

One recent proposal is that the large-scale dimensions of animacy and real-world object size organize our cognitive, neural, and perceptual object representations (Connolly et al., 2012; Julian, Ryan, & Epstein, 2016; Konkle & Caramazza, 2013; Konkle & Oliva, 2012a). Both of these dimensions are highly relevant to how we interact with objects; indeed, a critical function of the visual system is to identify whether something is animate, whether we can use it with our hands, or whether we should use it as a landmark. Accordingly, when we recognize an object, we very quickly know whether it is an animal (Thorpe, Fize, & Marlot, 1996) and how big it is in the real world (Konkle & Oliva, 2012b). At a neural level, these two dimensions jointly structure cortical responses to objects: Responses to big objects (e.g., couches, cars), small objects (e.g., cups, hammers) and animals (e.g., cats, horses) exhibit a tripartite organization that is mirrored across the lateral and the ventral surfaces of the cortex (Konkle & Caramazza, 2013). Perceptually, we have found that small objects look more like other small objects than they do like big objects, and vice versa (Long, Konkle, Cohen, & Alvarez, 2016). In particular, big and small objects differ in mid-level perceptual features, including texture and form (Long et al., 2016).

Here, we examine the hypothesis that animals and manmade objects also differ in mid-level perceptual features. We focus specifically on the distinction between animals versus artifacts (instead of all living versus all nonliving), as this distinction has recently

Citation: Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, *17*(6):20, 1–20, doi:10.1167/17.6.20.

doi: 10.1167/17.6.20 Received July 15, 2016; published June 27, 2017

been shown to reflect the organization of object-selective cortex (Konkle & Caramazza, 2013). Prior research hints that animates and artifacts may have different mid-level features. Simple line drawings of animals tend to have higher curvilinearity than line drawings of artifacts, lending some evidence to the intuition that animals are usually curvier than artifacts (Levin, Takarae, Miner, & Keil, 2001). However, line drawings may emphasize curvature differences, and there are likely other perceptual features that distinguish animates from artifacts (e.g., texture) that line drawings discard. In particular, recent work has found that natural scenes with animals—versus without—have different higher order texture statistics (Banno & Saiki, 2015).

As in previous work (Long et al., 2016), we used visual search performance as our index of perceptual similarity, as the speed of search is influenced by the similarity of the target to the distractors (e.g., Duncan & Humphreys, 1989; Levin et al., 2001; Long et al., 2016). Specifically, if animates and manmade objects are distinguishable in terms of features that influence visual search, then it should be easier to find an animal among objects than among other animals, and vice versa. To ask whether animates and inanimates differ in mid-level perceptual features, we then generated synthetic stimuli from pictures of animals and inanimate objects (see Figure 1; Freeman & Simoncelli, 2011), and measured visual search performance for these *texforms*. These texforms cannot be explicitly recognized at the basic level, but can be classified as animals versus artifacts. Given that texforms still carry semantic information about animacy, we directly assessed whether this visual search advantage arises during early stages of target individuation or semantic processing by recording the brain's electrophysiological activity. Specifically, we examined neural signatures of early target-distractor individuation (the N2pc component; Luck & Hillyard, 1994) and semantic processing (N400 component; for a review, see Kutas & Federmeier, 2011). Lastly, we investigated the degree to which perceived curvature is a key feature that distinguishes animates versus artifacts by creating several texforms that preserve or eliminate certain sets of visual features. In particular, testing these additional stimulus sets can reveal to what degree perceived curvature impacts whether texforms are classified as animals versus artifacts.

Across three visual search experiments, we found that search is more efficient when target and distractors differ in animacy, even when the stimuli are unrecognizable texforms. In addition, we found electrophysiological evidence that the mid-level features that distinguish animates versus inanimates can be used to individuate targets within the first few hundred milliseconds of processing, showing that no semantic processing is necessary to individuate animates from inanimates during visual search. Lastly, we find that the perceived curvature of a texform predicts whether it will be classified as animate or inanimate, and that curvature differences between animals and artifacts explain part of the mixed-animacy search advantage. Together, these results suggest that animates and artifacts differ in mid-level perceptual features, including, but not limited to, curvature, and that the visual system processes this difference within a few hundred milliseconds. Thus, we propose that there exist consistent perceptual differences between animates and artifacts that are available early on in processing to facilitate object detection, recognition, and classification.

## Experiment 1

In Experiment 1, we asked whether animates versus inanimates differ in terms of features that guide visual search. First we widely sampled a large set of images, and then we selected and controlled a subset of 120 images based on a number of criteria (see Stimuli). Then, we asked participants to find a target from one category (e.g., a deer) among distractors from the same category (i.e., animates) or among distractors from a different category (i.e., inanimate objects). If animals appear more similar to other animals, it should be harder to find the deer among other similar-looking animals than among dissimilar-looking objects. Thus, to the extent that there are consistent perceptual differences between animates and inanimates, we expect visual search to be more efficient when the target and distractors differed in animacy (e.g., a deer among objects; *mixed display*) versus when targets and distractors were from the same animate category (e.g., a deer among other animals; *uniform display*).

### Participants

Sixteen Harvard affiliates or students (age range: 18–35) were recruited and participated in this experiment. All studies were approved by the Institutional Review Board at Harvard University and designed in accordance with the principles expressed in the Declaration of Helsinki. All participants signed an informed consent prior to participation and reported normal or corrected-to-normal vision.

### Stimuli

We selected a wide range of images of animals and manmade objects using Google image search and

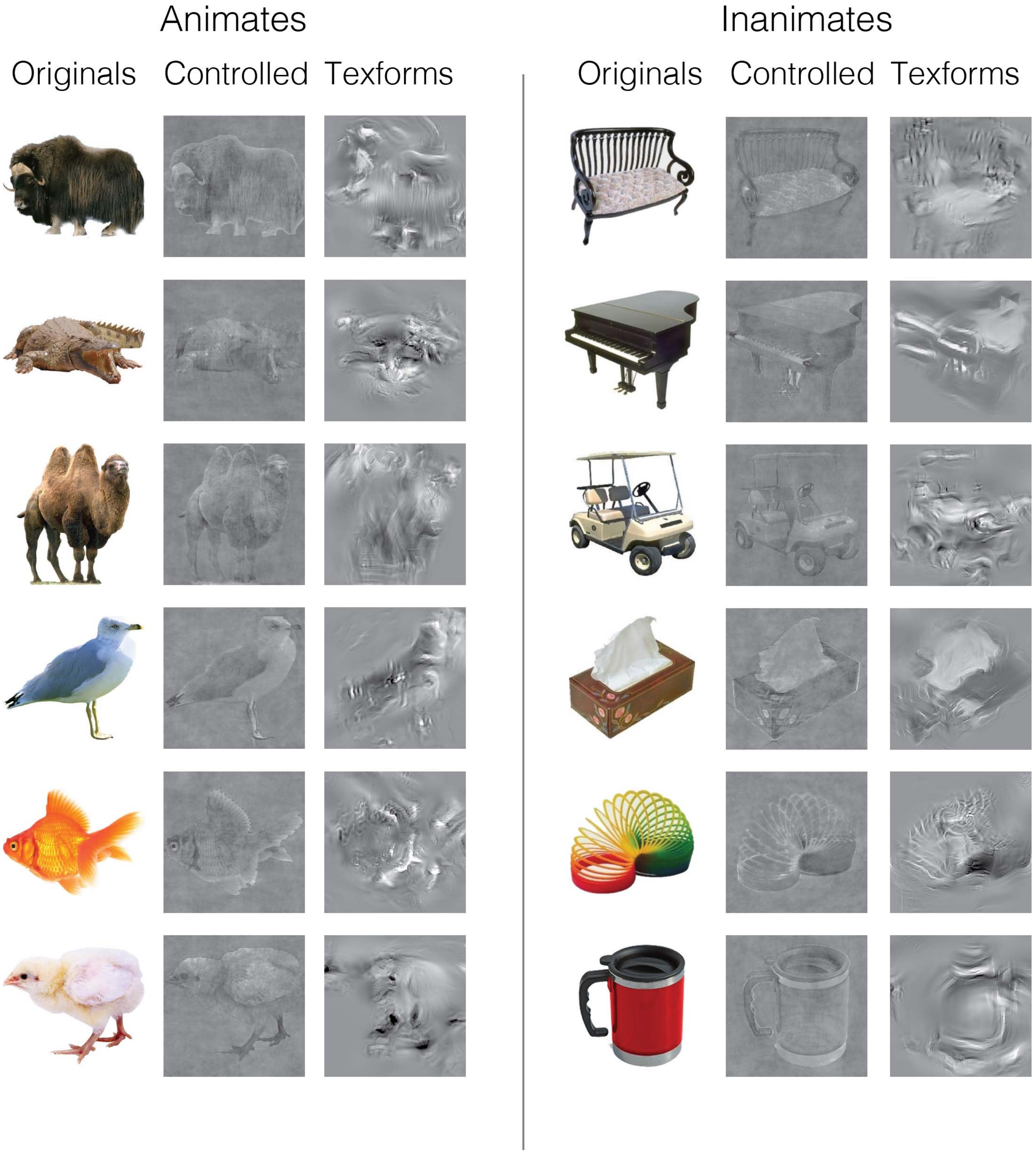

Figure 1. Example stimuli used in Experiments 1, 2, and 3. Original images are the color versions of the images used to generate both the controlled and texform versions. Controlled versions were created using the SHINE toolbox (Willenbockel et al., 2010); Texform stimuli were created using a texture-synthesis algorithm (Freeman & Simoncelli, 2011).

existing image databases (Konkle & Oliva, 2012a). In particular, we sampled big animals (e.g., gorillas, elk), big objects (e.g., cars, couches), small animals (e.g., cats, ants) and small objects (e.g., cups, shoes) so that we would be able to balance real-world size across animates versus inanimates. As big versus small objects have been shown to differ in their mid-level perceptual features (Long et al., 2016), it was important to control for this dimension. The final set included a wide range of animals, including mammals, aquatic animals, and insects. All manmade objects had a canonical orientation and also included a wide range of basic-level categories, including vehicles (i.e., sailboat, car), furniture (i.e., couch, bench), musical instruments (i.e., guitar set, piano), kitchen items (teapot, carafe, travel mug), and miscellaneous household items (stapler, slinky, Nintendo controller). All stimuli are available from the first author's website.

We then measured four low-level image properties of each image and equated them across categories. First, contour variance was measured by computing the standard deviation of the distance from the centroid of each object (Gonzalez, Woods, & Eddins, 2009) to each point on the object's contour, as previous research indicates this factor may influence visual search (Naber, Hilger, & Einhäuser, 2012). Second, object extent was taken as the ratio of the area of the object to its rectangular bounding box (Gonzalez et al., 2009). Third, image area was quantified (percentage of nonwhite pixels within a square frame) and fourth, aspect ratio (maximum height/maximum width in the picture plane) was measured. Finally, we also conducted Amazon Mechanical Turk experiments to obtain typicality ratings for each image on a 4-point Likert scale ($N = 108$).

We selected 60 images of inanimate and 60 images of animate objects (120 total). These final images were chosen so that animates versus inanimates did not differ on any of the above features (two-sample $t$ tests, all five $p$ values $> 0.14$). We also ensured that, within this larger stimulus set, big animals versus big objects as well as small animals versus small objects did not differ on any of the features listed above (two-sample $t$ tests, all $p$ values $> 0.08$). Lastly, these 120 selected objects and their backgrounds were also matched in terms of their intensity histograms (luminance and contrast) and power spectra (power at each orientation and spatial frequency) using the SHINE toolbox (Spectrum, Histogram, and Intensity Normalization Toolbox; Willenbockel et al., 2010). These images were set to an average luminance of 40.46 cd/m$^2$, presented on a lighter gray background (110.08 cd/m$^2$) to ensure they segmented from the background easily. Example stimuli are shown in Figure 1.

## Procedure

Participants performed a visual search task, in which they searched for a target object among a set of distractors (Figure 2A). On each trial, the exact target stimulus was previewed and presented centrally for 1000 ms. After 500 ms, a search display with either two or six items was presented in a circular array around fixation with three positions on either side of the vertical midline (see Figure 2A). The target was always present on the display, and the task was to locate the target as quickly as possible. Participants pressed the space bar as soon as they located the target, after which all items were replaced with numbers and participants clicked on the target's location using the mouse. This procedure enabled us to verify that participants had actually found the target. Distractors were either from the same-animate category (uniform-animacy displays) or the different-animate category (mixed-animacy displays) as the target (Figure 2B). Given that real-world size influences visual search efficiency (Long et al., 2016), distractors were always from the same size category. For example, on mixed-animacy displays, a big animal would appear among big object distractors (or vice versa), or a small animal would appear among small object distractors (or vice versa). Trial types were randomly intermixed throughout the session. Feedback was given after every trial, and accuracy was encouraged as incorrect responses resulted in a 5-s delay before the next trial could be initiated. There were 10 blocks of 96 trials, yielding 120 trials per condition (each combination of set size, target animacy, and display type). Reaction time and accuracy were recorded.

## Data preprocessing

Reaction times were excluded in which participants incorrectly identified the target or responded in less than 300 ms. In addition, we excluded reaction times that fell outside 3 $SD$s from the median deviation of the median (Rousseeuw & Croux, 1993), for each combination of set size, display type, and target animacy for each subject. On average, this excluded 14.41% of trials ($SD = 7.76\%$).

## Results

Our main question of interest is whether participants would find targets more efficiently on mixed-animacy displays (where distractors are from a different animate category than the target), than on uniform-animacy displays (where distractors are from the same animate

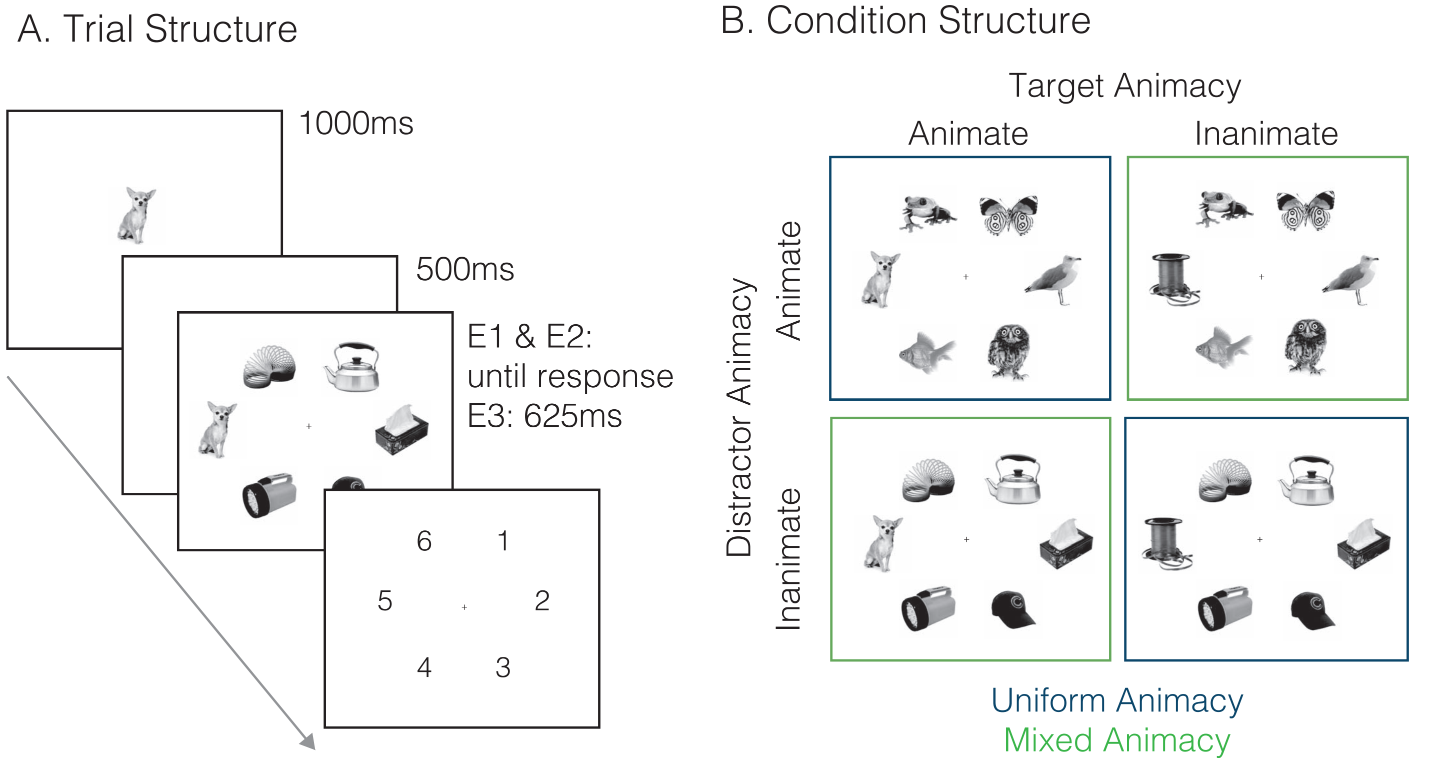

Figure 2. Trial and condition structure for Experiments 1 through 3. Actual stimuli are shown in Figure 1; grayscale stimuli are used here for illustration purposes only. (A) Visual search trial structure. First, the target was presented centrally for 1000 ms and then disappeared for 500 ms before reappearing among five distractor objects. In Experiments 1 and 2, these images remained on the screen until participants responded by pressing the space bar, after which the images were replaced by Xs (Experiments 1 and 2), and participants had to click on the correct location of the target. In Experiment 3, participants maintained fixation while the search display was shown for a limited amount of time (625 ms), after which participants were asked to indicate the location of the target by choosing a number on the keyboard (1 through 6). (B) Condition structure used in all experiments. We varied the animate category of the targets and distractors such that displays were either of uniform-animacy (blue/dark gray), where the targets and distractors were from the same animate category, or mixed-animacy (green/light gray) where the targets and distractors were from different animate categories.

category as the target). To address this question, we calculated the search efficiency for each of these conditions, based on the slope of the line relating reaction time to set size. The results indicate that search was about 28 ms/item more efficient on mixed relative to uniform displays, confirming our prediction: uniform slope: $M = 57.86$ ms/item, $SD = 26.27$ ms/item, mixed slope: $M = 29.84$ ms/item, $SD = 11.98$ ms/item, $t(15) = 6.37$, $p < 0.001$, Cohen's $d = 1.59$.

Next, we confirmed these results using a repeated-measures ANOVA with factors set size (2, 6), target category (animates, inanimates) and display type (mixed, uniform). As expected, we found that search was faster on mixed versus uniform displays: Figure 3A; main effect of display type, $F(1, 15) = 30.7$, $p < 0.001$, $\eta_p^2 = 0.67$. Critically, we found that search was also more efficient on mixed versus uniform displays, confirming our previous analysis using slopes: interaction of set size and display type, $F(1, 15) = 38.8$, $p < 0.001$, $\eta_p^2 = 0.72$.

We also examined whether there were any differences depending on whether the target was an animal versus an object. Overall, we found that search was slower when people were searching for an animal versus an object: main effect of target category, $F(1, 15) = 11.9$, $p < 0.01$, $\eta_p^2 = 0.44$; and that the mixed-animacy search advantage was also larger when the target was an animal versus an object; three-way interaction between set size, target category, and display type; $F(1, 15) = 7.64$, $p = 0.02$, $\eta_p^2 = 0.34$ (see Supplemental Figure 1A).

Overall these results imply that there are perceptual differences between animates and inanimates that observers used to more efficiently find targets on mixed-animacy displays.

## Experiment 2

In Experiment 1, when images were controlled for a variety of low-level perceptual features, visual search was more efficient when distractors differed from targets in animacy. These results may suggest that these

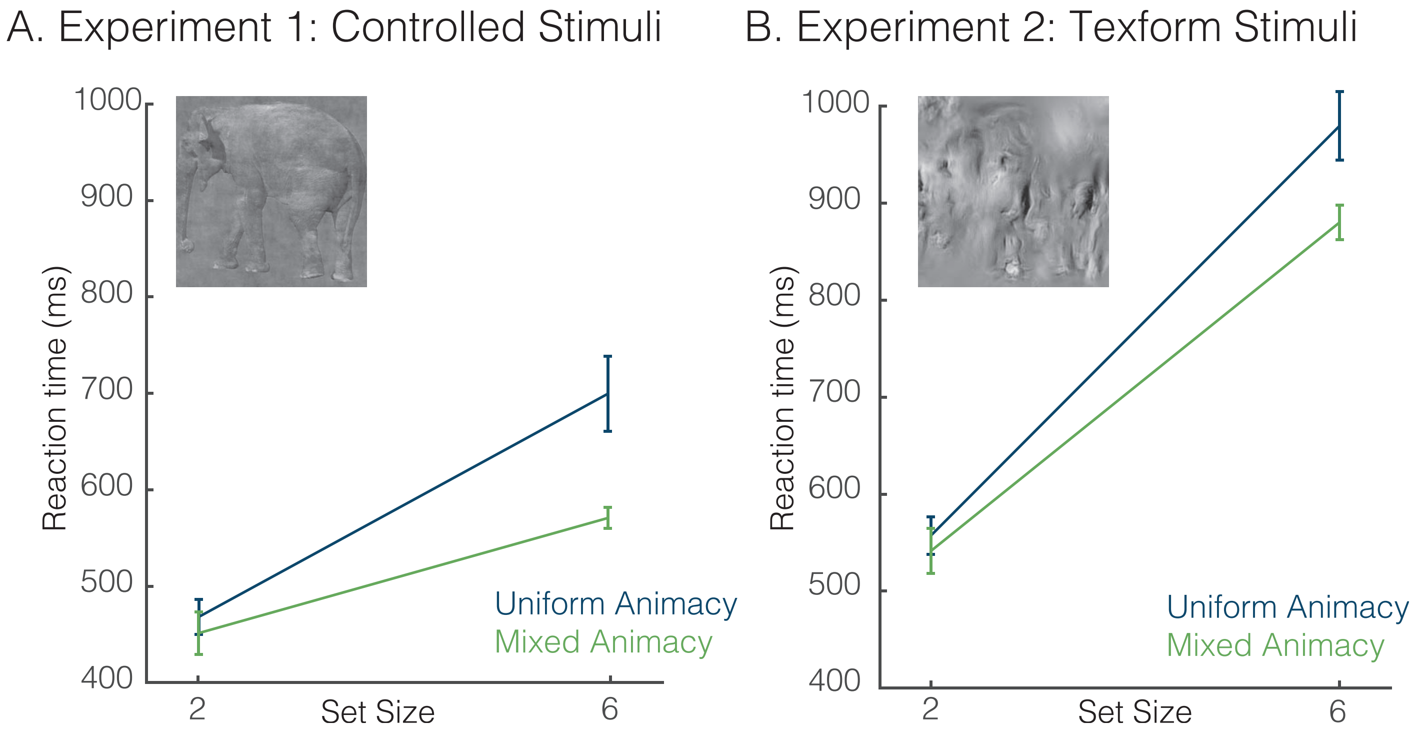

Figure 3. Results of Experiments 1 and 2. Search reaction times are plotted as a function of set size for uniform-animacy trials (when targets and distractors are of the same animacy) and mixed-animacy trials (when targets and distractors differ in animacy). Error bars indicate within-subjects confidence intervals (Morey, 2008).

animates versus inanimates differ in mid-level perceptual features that the visual system can extract for more efficient search. However, these images are also identifiable at the basic level and thus contain rich semantic information. Many researchers have proposed that semantic information can guide attention and eye movements during visual search (Becker, Pashler, & Lubin, 2007; Bonitz & Gordon, 2008; Hwang, Want, & Pomplun, 2011; Loftus & Mackworth, 1978; Underwood, Templeman, Lamming, & Foulsham, 2008; Wu, Wick, Pomplun, 2014; Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014). Thus, although other studies have found evidence against semantic guidance (De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks, Jr., & Hollingworth, 1999), it is important to address the possibility that on mixed-animacy displays, search was more efficient solely because distractors differed from the target in their semantic category. In Experiment 2, we asked directly if animates versus inanimates differ in mid-level perceptual features by creating unrecognizable versions of the same stimuli used in Experiment 1. Specifically, we created *texforms*, stimuli that preserve some texture and form information but obscure basic-level identification (Long et al., 2016). To create these texforms, we coerced white noise to have the same first- and second-order statistics as the original image (Freeman & Simoncelli, 2011).

If the gain in search efficiency observed in Experiment 1 is driven mostly by semantic differences that emerge as a product of intact basic-level recognition, then we should fail to find a gain in search efficiency with these mid-level texforms. If, however, the search advantage was driven by differences in mid-level feature differences between animates versus inanimates regardless of basic-level recognition, visual search should still be more efficient when targets and distractors differ in animacy.

## Participants

Sixteen Harvard affiliates or students (age range: 18–35) were recruited and participated in the visual search experiment and reported normal or normal-to-corrected vision. One hundred thirty five participants were recruited on Amazon Mechanical Turk and participated in the stimuli norming studies.

## Stimulus Generation

First, we generated texform versions of the 120 objects used in Experiment 1. To do so, we initialized Gaussian white noise images and iteratively adjusted them (using a variant of gradient descent) to conform to the modeled parameters of each original image (Freeman & Simoncelli, 2011; see Long et al., 2016, for a detailed description). The model output was saved after each iteration, and the final texforms were chosen from different model iterations for different stimuli to ensure that all texforms were unrecognizable at the basic level. These images were set to an average luminance of 54.83 cd/m$^2$, presented on a lighter gray background (110.08 cd/m$^2$) to ensure they segmented from the background easily.

## Stimulus norming

First, we verified that these texforms were unidentifiable at the basic level in an Amazon Mechanical Turk norming study ($N = 15$). Participants were shown all 120 images in a random order, one at a time, and asked to "guess what this could be." Responses were later coded for basic-level identity. If none of the observers guessed the correct basic-level identity of a texform, basic-level identification accuracy for this given texform would be 0%. Across all 120 texforms, the average basic-level identification accuracy was 1.83% ($SD = 5.6\%$).[1] Even when we coded accuracy liberally (e.g., accepting "hooved animal" as a correct answer for "deer"), basic-level identification accuracy across all texforms was only slightly higher 6.55% ($SD = 11.1\%$).

Secondly, to examine how well participants could classify these texforms as animate versus inanimates, we asked 120 observers to explicitly categorize these texforms as animals or manmade objects. All participants were told that they were viewing scrambled images. One group of 60 observers judged whether each texform generated was originally an animal or not, and another set of 60 observers judged whether each texform was generated from a manmade object or not. Each participant judged all 120 texforms in a single session. For each texform, we then calculated a classification score based on the proportion of correct classifications minus the proportion of incorrect classifications across subjects. This allowed us to control for any potential response biases. For example, consider a texform generated from a picture of a lion. When asked whether or not it was generated from a manmade object, this texform was classified as an "animal" 63% of the time, but when asked whether it was generated from a manmade object, it was also classified as a "manmade object" 20% of the time. Thus, subtracting the latter from the first score would yield an accuracy score of 0.43 for this particular texform. As a result, in this scoring system, a perfecty classified item would have a score of 1, a completely misclassified item would have a score of −1, and an item classified at chance would have a score of 0.

Overall, we found that texforms were classified as animate versus manmade objects at a rate above chance: $M = 0.19$, $SD = 0.20$, $t$ test against 0, $t(119) = 10.71$, $p < 0.001$; minimum classification = –0.34, maximum = 0.70). Critically, the degree to which texforms were identifiable at the basic level did not predict how well they were classified as animates versus inanimates (see Supplemental Figure 2, $r = 0.02$, $p = 0.76$). Thus, even though these texforms were unrecognizable at the basic level, they preserved mid-level features that allowed participants to classify them as animates virus inanimate at a rate above chance.

## Procedure

The procedure for the visual search experiment was the same as in Experiment 1, except that stimuli were mid-level texforms.

## Data preprocessing

Reaction times were excluded using the same procedure as Experiment 1. This excluded 14.23% of trials on average ($SD = 4.22\%$).

# Results

Our critical question of interest was whether we would find the same search efficiency advantage that we observed in Experiment 1, even when we replaced identifiable images with their mid-level texforms. We found that this was the case: Search was again more efficient when targets and distractors differed in animacy. Specifically, search slopes were almost 20 ms/item shallower on mixed-animacy trials $M = 84.63$ ms/item, $SD = 15.56$ ms/item, than on uniform-animacy trials, $M = 104.43$ ms/item, $SD = 24.8$ ms/item, $t(15) = 4.97$, $p < 0.001$ (see Figure 3B).

We confirmed these results using a repeated-measures ANOVA, with set size, target animacy, and congruency as factors. As expected, search was faster when there were fewer distractors: main effect of set size, $F(1, 15) = 390$, $p < 0.0001$, $\eta_p^2 = 0.963$); when the target was an inanimate texform: main effect of target animacy, $F(1, 15) = 49.3$, $p = 4.11\text{e-}06$, $\eta_p^2 = 0.767$); and when targets and distractor texforms differed in animacy: main effect of display type, $F(1, 15) = 34.9$, $p = 2.89\text{e-}05$, $\eta_p^2 = 0.699$). Critically, we again found an interaction between set size and display type, reflecting the difference in slopes we observed in the first analysis (interaction between set size and display type, $F(1, 15) = 25.6$, $p < 0.001$, $\eta_p^2 = 0.631$). Unlike Experiment 1, we found that this gain in search efficiency was not modulated as a function of whether the target was an animal or an object: no three-way interaction between set size, display type, and target animacy, $F(1, 15) = 0.156$, $p = 0.698$, $\eta_p^2 = 0.01$; (see Supplemental Figure 1B).

## Experiment 3

The results of Experiments 1 and 2 suggest that animals and artifacts have different mid-level perceptual features that can be used to efficiently guide target selection and individuation. However, in Experiment 2 the stimuli were still classifiable as animates and inanimates, despite the fact that they were not identifiable at the basic level. Thus, some semantic information was still available to the participants, rendering it difficult to conclusively pinpoint the processing stage at which these effects arose. Indeed, behavioral responses reflect the product of multiple processing stages, including low-level sensory processing, target individuation, identification, response selection, and semantic activation. We thus used electrophysiology to directly test at what processing stage the mixed-animacy search advantage for texforms first originates.

We focused our analyses on a well-established neural marker of target-distractor individuation during visual search—the N2pc component of the event-related potential (Luck & Hillyard, 1994; Woodman & Luck, 2003). This component is associated with the successful individuation of a target from its distractors occurring prior to target identification and usually occurs at about 180–300 ms after search display onset (e.g., Mazza, Turatto, Umiltà, & Eimer, 2007). Importantly, the N2pc is measured with respect to the target location: Specifically, it reflects a difference in activity over posterior electrode sites that are contralateral versus ipsilateral to the target location. As a result, any difference observed in the N2pc component reflects the differential processing of the relationship between the targets and distractors, rather than perceptual properties of the display.

In general, a larger N2pc is elicited when targets are easily individuated relative to when they are not. Pop-out search displays tend to elicit an earlier and more consistent N2pc than demanding search displays (Dowdall, Luczak, & Tata, 2012). In addition, when distractors share features with the target, they can also capture attention and elicit their own N2pc, effectively cancelling out the N2pc observed in response to the target (Luck & Hillyard, 1994). Thus, if differences in mid-level features are used during the process of target individuation, we expect to observe a larger N2pc on mixed-animacy trials, when search is relatively efficient, and an attenuated or delayed N2pc on uniform-animacy trials, when distractors are perceptually more similar to the target and may themselves elicit a weak N2pc.

However, though texforms cannot be identified at the basic level (e.g., as a "car"), observers can still tell whether they refer to "animals" or "artifacts." Thus, a second possibility is that this broad level of semantic information is activated during visual search and generates the mixed-animacy search advantage we observed in Experiments 1 and 2. On this account, observers individuate targets from distractors equally well on mixed animacy and uniform animacy displays. However, on uniform-animacy displays, semantic interference from the same-animacy distractors slows observer's ability to verify that the target is actually the image they were looking for. If this is the case, we might expect to see a difference between uniform and mixed-animacy trials originating *only* later in processing (>300 ms), during components typically associated with semantic expectations, for example the N400 (for a review, see Kutas & Federmeier, 2011). Though the N400 has most frequently been used to study semantic violations during language processing, it also arises in response to pictures that are semantically unrelated to a previously seen picture (Barret & Rugg, 1990; McPherson & Holcomb, 1999). Furthermore, this same component also arises in response to semantically incongruous pictures of actions (e.g., someone cutting bread with a large saw; Proverbio & Riva, 2009) and pictures of scenes that contain semantically incongruent objects (e.g., a file folder in a dishwasher in a kitchen; Võ & Wolfe, 2013, see also Ganis & Kutas, 2003; Mudrik, Lamy, & Deouell, 2010).

Thus, if the mixed-animacy search advantage is driven by a difference in distractor rejection on the basis of *semantic* category membership, then we should observe a difference between mixed and uniform-animacy displays during the N400 time interval. Specifically, we should observe a larger N400 on mixed-animacy (vs. uniform-animacy) displays, where there is a mismatch in the semantic category membership between the targets and distractors.

An intermediate possibility is that semantic differences contribute to the difference in search performance on mixed versus uniform animacy displays, but that mid-level feature differences between animates and inanimates still help observers individuate targets on mixed-animacy displays. If this is the case, then we should observe both a greater N2pc component and a greater N400 component on mixed-animacy displays.

Finally, though we did our best to control these differences, the mixed-animacy search advantage could instead arise from simple low-level feature differences between our image sets. To rule out this alternative, we also examined how uniform versus mixed-animacy displays affected both the P1 and the N1 components, two well-studied components of the visual-evoked potential that index early visual processing (Mangun & Hillyard, 1991). The P1 is known to be sensitive to low-level feature differences, including luminance and contrast (Johannes, Münte, Heinze, & Mangun, 1995). Furthermore, some previous work has found greater N1 responses to line drawings of animals versus artifacts, possibly suggesting that the N1 has some sensitivity to perceptual features reflected in the animacy distinction (e.g., Kiefer, 2001; Proverbio et al., 2007; but see Antal et al., 2000). Thus, if mixed versus uniform displays have different basic, low-level properties, they should elicit different amplitudes in the P1-N1 complex of the event-related potential.

# Methods

## Participants

Eighteen Harvard students or affiliates participated in the experiment after giving informed consent. Data from two participants were excluded from analysis as they had less than 100 usable trials per condition due to artifacts in the EEG (e.g., muscle movements artifacts, eye movements, or eye-blinks). Sixteen participants (age range, 18–35 years) were included in the final sample, all of whom had normal or corrected-to-normal vision and were right-handed.

## Procedure

The procedure was a version of Experiment 2 that was modified so that we could record the brain's electrophysiological activity. Participants were tested in a sound-attenuated and electrically shielded booth that was dimly lit throughout the experiment. Participants were seated 57 cm in front of a 13-in CRT monitor (1024 × 768 pixels; refresh rate, 60 Hz). Participants were instructed to maintain their gaze at a black fixation cross in the center of the screen (see Figure 2A). Trials in which the participant failed to keep fixation were excluded from analysis, though fixation was monitored online by the experimenter throughout the session and participants were given verbal feedback after those trials in which they broke fixation.

The search display remained on the screen for 625 ms, after which numbers corresponding to their location replaced all images. The participants' task was to localize the target by pressing a key on the number pad of a keyboard that corresponded to the location of the target. Accuracy was emphasized over speed. In total, there were 20 practice trials and 960 test trials, counterbalanced such that each image appeared equally often with different category distractors and on either side of the screen (left/right). Feedback was given for 500 ms after each trial, after which the next trial started after a random interval between 500 and 1500 ms.

In addition, before the EEG experiment, all participants completed 106 practice trials. To gradually familiarize participants with the rapid presentation of the stimuli during the EEG experiment, we used the Quest algorithm to adjust the display time separately for each participant and condition during this practice session (Watson & Pelli, 1983).

## Behavioral analysis

Practice trials were excluded from analysis. Given that behavioral responses were not speeded, the analysis focused on accuracy and not RT in this experiment. Participants' average accuracies across our four main conditions (target category × display type) were submitted to a repeated-measures ANOVA.

## Electrophysiological recordings and analysis

EEG was recorded continuously from 32 Ag/AgCI electrodes mounted in an elastic cap and amplified by an ActiCHamp amplifier (BrainVision). Electrodes were arranged according to the 10–20 system. The horizontal electrooculogram was acquired using a bipolar pair of electrodes positioned at the external ocular canthi, and the vertical electrooculogram was measured at electrode FP1, located above the left eye. All scalp electrodes were referenced to an electrode on the left mastoid, and were digitized at a rate of 500 Hz. Continuous EEG data were filtered with a bandpass of 0.05–112.5 Hz offline.

The EEG analysis was performed for correct trials only. Individual trials were rejected when they were contaminated with ocular artifacts or muscle movement artifacts. Muscle movement artifacts were detected if the maximum voltage on any given channel exceeded 100 microvolts; or if there was a 100-microvolt change in voltage within a 200 ms window (calculated every 50 ms). Blinks were detected using a similar semiautomated procedure, in which differences between minimum and maximum voltages were compared with a threshold value over a 200-ms time window; threshold values were determined by visual inspection for each subject individually (Störmer et al., 2013). Eye movements were detected by using the horizontal electro-oculogram, and threshold values were also determined by visual inspection. Artifact-free data were digitally re-referenced to the average of the left and right mastoid. Statistical analyses were performed on the unfiltered data, but ERPs were digitally low-pass filtered (–3dB at cutoff frequency 30 Hz) for illustration purposes.

To measure the P1, N1, and N400, event-related potentials (ERPs) time-locked to the search display were averaged separately for trial type (mixed, uniform). The start of the P1 (70–90 ms), N1 (140–160 ms), and N400 (400–600 ms) components were determined based on time points identified in previous literature and by visual inspection of the average waveform across all conditions. Then, we examined average activity for uniform-animacy and mixed-animacy displays over several occipital electrode sites (O1, O2, Oz, P03, P04, P0z) for the P1 and N1, and over a group of central-electrode sites for the N400 (Cz, C3, C4, CP1, CP2, CP5, CP6).

To measure the lateralized N2pc component, ERPs time-locked to the search display were averaged separately for trial type (mixed, uniform) and target location (left or right side of the display; this included all three target locations for each side). Next, ERPs were collapsed across left and right target location and hemisphere of recording (left, right) to reveal waveforms recorded ipsilaterally and contralaterally with respect to the target location. ERP waveforms were measured for each participant and condition (mixed, uniform) with respect to a 100-ms prestimulus period at lateralized posterior electrode sites (P03/P04, P07/P08, and P7/P8). The N2pc was measured in two steps. First, to examine the amplitude of the N2pc component across the two conditions, the mean amplitude of the ipsilateral and contralateral waveforms was measured between 180–300 ms postsearch display onset (cf. Mazza et al., 2007). To test the significance levels of the N2pc, these mean amplitudes were subjected to an ANOVA with factors of electrode location (ipsilateral, contralateral) and display type (uniform, mixed).

Second, to identify at what point in time the N2pc emerged for the two conditions, we measured the difference between contralateral and ipsilateral waveform in successive 20-ms time bins beginning at 180 ms after display onset (i.e., immediately after the N1 component). The mean amplitude of each averaged time bin was tested with a *t* test to identify the time interval at which the N2pc became present. The onset time for the N2pc was taken as the first significant 20 ms that was followed by at least two or more successive time intervals.

## Behavioral results

Our main question of interest was whether we would again see a search advantage when texforms were presented on mixed-animacy relative to uniform-animacy displays. Consistent with Experiment 2, we found that participants were more accurate at finding texforms on mixed-animacy displays (see Figure 4A): repeated-measures ANOVA, main effect of display type, $F(1, 15) = 115$, $p = 0.0001$, $\eta_p^2 = 0.88$. In addition, we also found that participants had higher performance when searching for inanimates relative to animates: main effect of target animacy, $F(1, 15) = 82.2$, $p < 0.0001$, $\eta_p^2 = 0.85$. However, the mixed-animacy search advantage was again not greater when the target was an animate versus an inanimate texform, though there was a trend in this direction: no interaction between target animacy and display type, $F(1, 15) = 3.62$, $p = 0.08$, $\eta_p^2 = 0.195$.

## EEG results

### P1 and N1

We found no evidence that mixed-animacy and uniform-animacy displays generated low-level processing differences captured by the P1 component, occurring around 70–90 ms after display onset. The average voltages elicited by mixed and uniform-animacy displays were not statistically distinguishable from one another: $M_{diff} = 0.11$ μV, $SD_{diff} = 0.61$ μV, $t(15) = 0.71$, $p = 0.487$.[2]

Similarly, uniform and mixed-animacy displays did not generate differences in the N1 components: $M_{diff} = 0.19$ μV, $SD_{diff} = 0.95$ μV, $t(15) = 0.81$, $p = 0.43$. Thus, these results imply that the texforms did not generate differences in low-level sensory processing that the P1 or N1 components are sensitive to (see Supplemental Figure 3A).

### N2pc

We found that the electrophysiological activity over the hemisphere contralateral to the target location diverged from the activity over the hemisphere ipsilateral to the target location, providing clear evidence for the N2pc, 180–300 ms, $t(15) = 2.87$, $p = 0.01$, Cohen's $d = 0.72$. The magnitude of the N2pc component was sensitive to display type: On trials with mixed-animacy displays, the N2pc was larger relative to trials with uniform-animacy displays (see Figure 4B and 4C): mixed, $M = -0.43$ μV, $SD = 0.46$ μV; uniform, $M = -0.06$ μV, $SD = 0.42$ μV; interaction between N2pc and display type, $F(1, 15) = 6.74$, $p = 0.02$, $\eta_p^2 = 0.31$,[3] consistent with the hypothesis that mid-level perceptual features help to quickly individuate the target from the distractors.

Figure 4D shows the contralateral-minus-ipsilateral difference waveform separately for uniform and mixed trials. To quantify at what point in time the N2pc emerged, we examined this difference waveform in

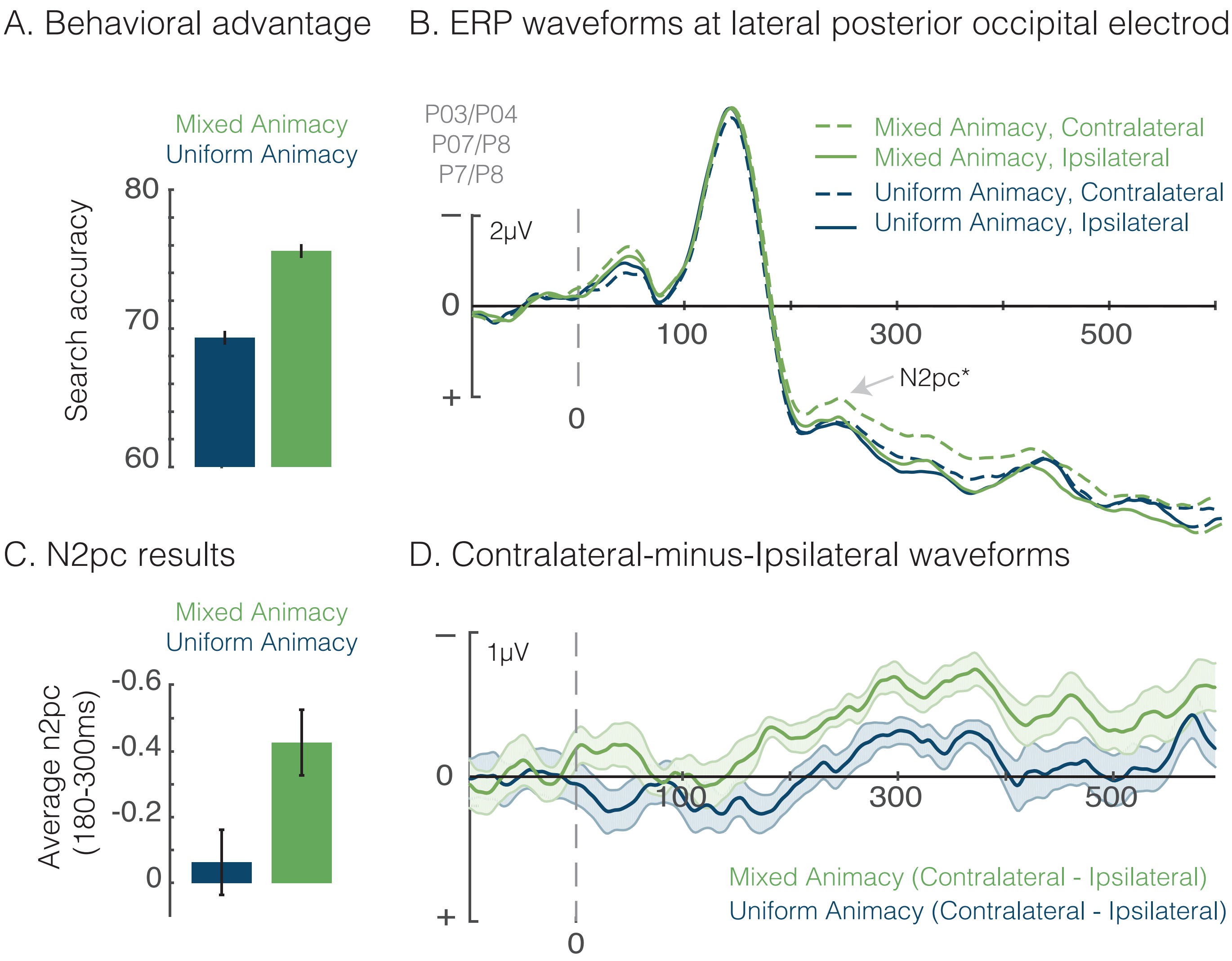

Figure 4. (A) Search accuracy as a function of display type. (B) ERP waveform over lateral parietal-occipital scalp sites, time-locked to the onset of the search display. ERP waveforms are plotted separately for mixed and uniform displays, as well as when the target was localized contralateral versus ipsilateral to the electrodes. (C) Mean amplitude of the N2pc component for each display type between 180–300 ms. (D) Contralateral-minus-Ipsilateral difference waveforms, plotted separately for mixed and uniform-animacy displays. Error bars represent within-subjects standard error of the mean (Morey, 2008).

successive 20 ms time bins separately for each condition from 180 ms through 600 ms after display onset. For the mixed displays, the N2pc became significant at 220 ms and lasted until 380 ms. In contrast, for uniform displays, the N2pc became significant 60 ms later, at 280 ms post display onset, and lasted until 320 ms.

#### *Late processing component (N400)*

Although these texforms were unrecognizable at the basic level, they could still be classified as animals versus artifacts. Participants in this study were not asked to identify or categorize the texforms during the search task, but it is possible that semantic representations are automatically activated by the texforms, and that semantic congruency between the targets and distractors played a role in the behavioral search advantage on mixed-animacy displays. If this were the case, then this semantic congruency effect should be reflected in the N400, a late-occurring processing component sensitive to semantics (e.g., Võ & Wolfe, 2013). Thus, we examined whether uniform versus mixed-animacy displays generated different neural responses in the N400 time range (i.e., 400–600 ms). In contrast to the N2pc, the N400 is a nonlateralized component typically observed over centro-parietal scalp areas (Kutas & Federmeier, 2011); thus, we performed these analyses using ERP waveforms averaged over the location of the target (i.e., nonlateralized EEG activity). We chose this time window based on prior literature (Kutas & Federmeier, 2011).

We found that there was no difference in the activity generated at central electrodes sites between 400 and 600 ms postdisplay onset for uniform versus mixed conditions: $M_{diff} = 0.04$ μV, $SD_{diff} = 0.51$ μV, $t(15) = 0.33$ $p = 0.746$ (see Supplemental Figure 3B). This same pattern held when we examined the occipital electrodes: $M_{diff} = -0.11$ μV, $SD_{diff} = 0.59$ μV, $t(15) = -0.72$, $p = 0.481$ (see Supplemental Figure 3A). Thus, mixed and uniform-animacy displays did not generate any differences in the N400 component (for full scalp topographies between 350 and 600 ms, please see Supplemental Figure 4).

### Early frontal electrode analyses

The prior analyses suggest that mixed and uniform displays do not generate differences in semantic processing that are indexed by late-occurring components sensitive to semantic violations. However, another possibility is that the semantic congruency between the targets and distractors is instead extracted very early in processing, as some studies have suggested (e.g., Thorpe et al., 1996). If so, such early differences between mixed versus uniform-animacy displays would be expected to be present at frontal electrode sites (Thorpe et al., 1996). Furthermore, such an early-occurring difference at frontal electrodes could complicate our interpretation of the similarly timed N2pc differences. Thus, we examined whether responses at frontal electrodes (Fz, F3, F4, FC5, FC6, FC1, and FC2) differed in response to uniform versus mixed conditions at any point during the trial. We computed paired $t$ tests between the average voltages to uniform versus mixed trials over 20 ms bins, from 100 ms through 600 ms after display onset. No two successive time intervals reached significance (see Supplemental Figure 3C).

Overall, we did not find any evidence for components that were sensitive to the semantic congruency between the targets and distractors. Thus, we conclude that it is unlikely that any early differences in semantic processing could have influenced the N2pc results.

## Discussion

Overall, these results suggest that the behavioral search advantage observed in both Experiments 2 and 3 was due to a faster and more efficient individuation of the target from the distractors on mixed relative to uniform animacy trials. Participants were more likely to locate the target during the first few hundred milliseconds on mixed-animacy versus uniform-animacy trials, leading to the larger N2pc observed on mixed-animacy trials. In addition, we found no evidence that this behavioral advantage originated from either differences in low-level processing indexed by the P1/N1 components, or from semantic processing indexed by early frontal or late-occurring central or occipital components. Thus, these electrophysiological results extend and confirm our behavioral results from Experiment 2 by pinpointing the locus of the behavioral effect to target individuation.

## Experiment 4

In Experiments 1 through 3, we found evidence that animals and artifacts differ in early perceptual features that can be used to rapidly individuate targets from distractors during visual search. In Experiment 4, we sought to identify one of these perceptual features: In particular, we examine the degree to which perceived curvature is a key feature that distinguishes animals versus manmade objects and explains the mixed-animacy search advantage. To do so, we first examined the features that observers use to classify texforms as animates versus artifacts. Then, we reanalyzed the behavioral data from Experiment 2 as a function of the difference in perceived curvature between the target and distractor texforms.

First, we examined if perceived curvature predicts the ability to classify texforms as animates versus inanimates. Then, we examined how preserving versus degrading the curvature information in texforms impacts how well they are classified by their animacy. When images are reduced to texforms, they can still be classified as animals versus manmade objects, even though they are unrecognizable at the basic level (see Experiment 2). We examined how animate/inanimate classification is impacted when we simply flipped the vertical orientation of the texforms or reversed their contrast (Balas, 2012), two manipulations that we hypothesized should not impact classification performance as they do not alter perceived curvature.[4] Next, we examined animate classification performance for texforms that only have power at vertical and horizontal spatial frequencies. As this manipulation eliminates curvature differences between animate and manmade object texforms, we expected classification performance to decrease. Lastly, we examined classification performance for texforms that lack higher order cross-correlation statistics that capture all kinds of curvature information. Previous work has found that these higher order statistics (cross position, cross scale, and cross orientation correlations of V1-like responses) explain part of perceptual sensitivity in natural texture perception (Freeman et al., 2013; Okazawa, Tajima, & Komatsu, 2015). More generally, this "model-lesioning" approach has been successfully used to study the relevant features in texture perception (Balas, 2006). As a result, "texforms" that are generated without cross-correlation statistics should lack the critical features that allow observers to distinguish animates from manmade objects. Thus, we expected classification performance to fall to chance when these cross-correlation statistics were removed from the texture synthesis procedure.

Finally, we examined the degree to which curvature differences between animals and manmade objects explain the fact that participants searched more efficiently when targets and distractors differed in animacy. Specifically, we reanalyzed search reaction times from Experiment 2 as a function of the perceived curvature difference between targets and distractors. If the mixed-animacy search advantage is driven entirely by curvature differences between targets and distractors, then there should be no mixed-animacy search advantage when targets and distractors differ little in perceived curvature. Alternatively, curvature differences between targets and distractors may explain part of the mixed-animacy search advantage. If so, then we should find more efficient search when targets and distractors differ more in curvature, but the mixed-animacy search advantage should still persist even when the target and distractors differ little in curvature.

## Methods

### Participants

Participants were recruited on Amazon Mechanical Turk. Sixty observers were recruited for each animate classification study ($N = 30$ per question), and 30 observers were recruited for the curvature ratings.

### Procedure

As in the norming studies reported for Experiment 2, two sets of observers judged "whether the thing in the scrambled image was an animal or not" or "whether the thing in the scrambled image was a manmade object or not." In addition, another set of observers judged how "boxy" or "curvy" each of the 120 texforms were on a 5-point scale (1 = Very Curvy, 2 = Somewhat Curvy, 3 = Neither Curvy nor Boxy, 4 = Somewhat Boxy, or 5 = Very Boxy).

### Stimuli

Using custom image scripts and modifications of the texture synthesis model, we created four different versions of the stimulus set used in Experiment 2. First, we took the 120 texforms used in Experiments 2 through 3 and created three modified versions. Second, we created inverted versions of the texforms by simply flipping these images upside down and created contrast-negated versions using custom scripts in Matlab (MATLAB 2015a, MathWorks, Natick, MA).

Third, we created texforms that retained energy primarily at horizontal and vertical orientations. This was done by taking a discrete Fourier transform of each image (Matlab 2015a, function fft2), retaining power in the first and last ~1% of pixels of the Fourier representation of a 640 × 640 pixel image and setting power at all other locations to zero (see Supplemental Figure 5). The inverse of this modified Fourier transform yielded a modified version of each image with power primarily at orientations near horizontal and vertical. We verified that these modified images had the greatest power at horizontal (i.e., 180°) and vertical (i.e., 90°) orientations by examining the rotationally averaged spectrum of each image.

Finally, we created a new synthesized version of these images by modifying the parameters of the texture synthesis algorithm. Specifically, we lesioned (i.e., eliminated) higher order correlations of V1-like responses (linear and energy filters) across positions, orientations, and spatial scales from the texture synthesis algorithm (Freeman et al., 2013). In other words, these correlations were not adjusted during the image synthesis procedure and the "texforms" instead were generated without these parameters.

All image sets and a link to the texform generation code are freely available from the first author's website (see Figure 5 for examples).

### Visual search curvature analysis

Here, we reanalyzed the behavioral data from Experiment 2 as a function of the perceived curvature of the targets and distractor texforms. To do so, for each trial and participant we computed a score of the perceived curvature difference between targets and distractors. To do so, we averaged the perceived curvature ratings for all distractors on a given trial, subtracted them from the perceived curvature score for the target, and took the absolute value of this score. We then sorted all trials, irrespective of condition, on the basis of this curvature difference score between targets and distractors. For each subject, trials where the curvature difference score was greater than the median difference were labeled "high curvature difference trials" and trials with scores lower than the median difference were labeled "low curvature difference trials." We then performed targeted $t$ tests over the search slopes (as in Experiment 2) separately for high

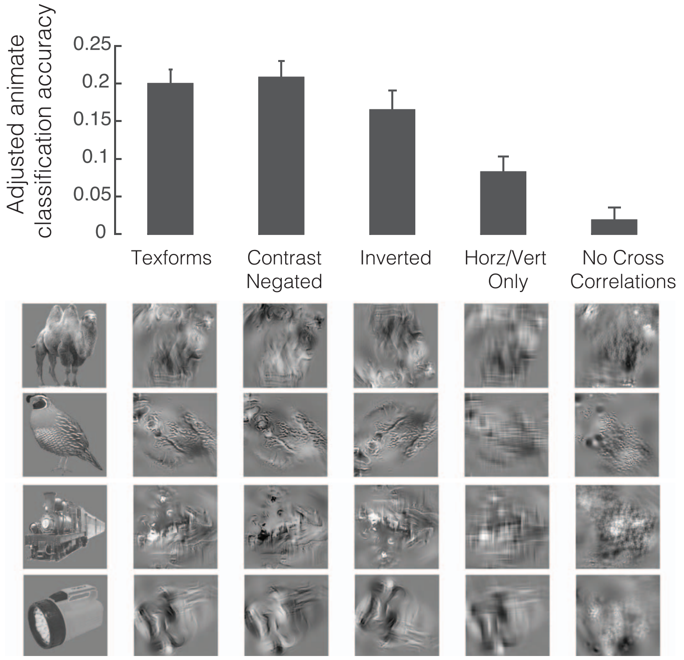

Figure 5. (Upper panel) Adjusted animate classification accuracy for each of the five modified image sets. Zero reflects chance classification. Error bars reflect standard error of the mean over individual items. (Lower panel) Four example items from each image set.

curvature and low curvature difference trials. Thus, if curvature differences drive the mixed-animacy search advantage, we should expect a smaller difference in search efficiency on mixed versus uniform trials on low curvature difference trials.

## Analysis and results

Overall, we found that texforms generated from animals versus artifacts differed in perceived curvature, such that texforms generated from animals were perceived as more curvy than texforms generated from artifacts: Animal texforms, $M = 2.33$; Object texforms, $M = 2.77$, $t(118) = 6.36$, $p < 0.001$. While texforms generated from artifacts spanned the entire range of curvature ratings (Range: 1.57–4.23) texforms generated from animals were all rated as relatively curvy (Range: 1.67–2.87; $F$ test for unequal variances, $F = 4.74$; $p < 0.001$). Lastly, the degree to which an animal texform was perceived as curvy predicted how well it would be classified as animate ($r = 0.32$, $p = 0.01$), and the degree to which an object texform was perceived as boxy predicted how well it would be classified as an artifact ($r = 0.57$, $p < 0.001$). Thus, perceived curvature appears to be one perceptual feature that observers use to classify these texforms as animals versus artifacts.

We next examined classification performance when we preserved or decreased curvature information in the texforms. As expected, flipping the texforms on their vertical axis or negating their contrast did not change observers' ability to classify them as animals or manmade objects relative to unmodified texforms, $t(119) = 1.86$, $p = 0.07$, contrast negated, $t(119) = -0.43$, $p = 0.67$. Thus, when curvature differences are maintained in the texforms, participants can classify these stimuli as animate/inanimate.

Next, we found that participants' ability to classify these images as animate/inanimate was much worse for texforms that only had power at horizontal and vertical spatial frequencies compared to unmodified texforms, paired $t$ test, $t(119) = 4.63$, $p < 0.001$. While observers could still classify object texforms as manmade objects, $M = 0.21$, $t(59) = 8.63$, $p < 0.001$, they could no longer identify which texforms referred to animals, $M = -0.04$, $t(59) = -1.78$, $p = 0.08$. Lastly, we examined classification performance when we removed all cross-correlation statistics from the model that would carry information about rectilinearity and curvilinearity. When we eliminated all cross correlations, classification accuracy dropped to chance for images generated from either objects or from animals: animals, $M = 0.004$, $t(59) = 0.21$, $p = 0.84$; objects, $M = 0.032$, $t(59) = 1.22$, $p = 0.23$ (Figure 5). Overall, these results suggest that the combination of multiple cross-correlation statistics in the model contribute to the creation of perceptual features, including curvature, that allow observers to classify them as animate/inanimate.

If curvature is an important feature for distinguishing animates from artifacts, then perceived curvature differences between targets and distractors may contribute to the mixed-animacy search advantage. We reanalyzed the reaction time data from Experiment 2 as a function of the perceived curvature between the targets and distractors of the texforms. Overall, we found that the difference in perceived curvature between targets and distractors impacted visual search efficiency, but did not entirely explain the mixed-animacy search advantage. We compared performance on trials with high curvature differences between target and distractors to trials with a low curvature difference. Search was more efficient when there was a difference in perceived curvature between targets and distractors, which was reflected in overall shallower search slopes for high-curvature difference trials, regardless of the animacy of the targets and distractors: high-curvature difference, $M = 86.81$ ms/item; low curvature difference, $M = 102.5$ ms/item, $t(15) = -6.34$, $p < 0.001$. Nonetheless, search was still more efficient on mixed-animacy relative to uniform-animacy trials, both when there was a low difference in perceived curvature between targets and distractors [Uniform, $M = 109.90$ ms/item; Mixed, $M = 89.71$ ms/item, $t(15) = 5.23$, $p <$

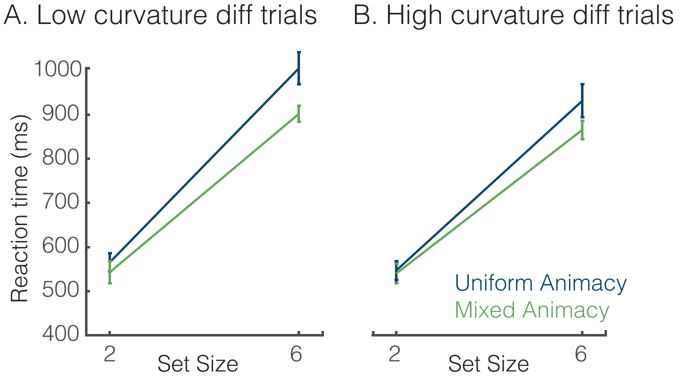

Figure 6. Reaction-time data from Experiment 2, reanalyzed as a function of the curvature difference between the targets and the distractors. (A) Average search reaction times for uniform and mixed-animacy trials (for texforms) when there was a relatively low difference in perceived curvature between targets and distractors. (B) Average reaction times for uniform and mixed-animacy trials when there was a high difference in perceived curvature between targets and distractors. Error bars represent within-subjects 95% confidence intervals (Morey, 2008).

0.001 (see Figure 6A)] and when there was a high difference in perceived curvature between targets and distractors [Uniform, $M = 96.09$ ms/item; Mixed, $M = 81.35$ ms/item, $t(15) = 2.71$, $p = 0.016$ (Figure 6B)].

Thus, the mixed-animacy search advantage held regardless of the difference in perceived curvature between targets and distractors, yet search was most efficient when targets and distractors differed in curvature and animacy: Search slopes for high-curvature difference mixed trials versus low-curvature difference mixed trials, $t(15) = 3.06$, $p < 0.01$. These results suggest that a difference in perceived curvature between animals and manmade objects impacts visual search efficiency, but does not entirely explain the mixed-animacy search advantage observed in Experiments 1 through 3.

## General discussion

Here, we found that visual search was more efficient when a target differed from distractors in animacy. For example, participants more easily found animals among inanimate objects than among other animals. We found this behavioral search advantage when targets and distractors were controlled for a wide range of low-level features (Experiment 1) as well as when these same stimuli were reduced to texforms—texturized stimuli that loosely preserve an object's texture and form but cannot be explicitly recognized at the basic level (Experiment 2). We then confirmed that this behavioral advantage was reflected in an early ERP signature of target-distractor individuation—the N2pc component (Experiment 3) and not in components that process low-level sensory differences (i.e., the P1 and N1) or semantic congruency (i.e., the N400). Lastly, we found that perceived curvature is a key feature that distinguishes animates versus artifacts (Experiment 4).

Taken together, these findings suggest that the mid-level perceptual differences between animates and manmade objects, including curvature, are available to the attentional system within ~220 ms. Thus, these mid-level perceptual features could potentially serve as early cues to animacy in the visual system. Though most models of object representation assume that higher level information is accessed after basic-level recognition has been achieved, the present findings suggest that cues to animacy are available without explicit recognition, and could enable the system to make a first guess at whether something is animate or an artifact.

### What are the relevant features?

In the present study we first focused on determining that animates and artifacts differed in terms of their mid-level perceptual features and then examined the contribution of one candidate feature—perceived curvature. While the perceived curvature of man-made objects ranged from very curvy to very boxy, all animals tended to be perceived as relatively curvy. Thus, in perceived curvature space, animals cluster within artifacts. Accordingly, we found that the degree to which a texform was perceived as "boxy" predicted how well it was classified as an artifact (versus an animal). In addition, when texforms no longer contained reliable curvature information, observers' ability to classify these texforms as animate/inanimate decreased. In particular, when we eliminated cross-correlations of V1-like responses from texforms, animate classification performance fell to chance, suggesting that these statistics carry the key mid-level features that distinguish animates versus artifacts.

However, perceived curvature did not predict all of the variance in observers' ability to classify these texforms by their animacy nor entirely explain the mixed-animacy search advantage. Thus, these results also suggest that there are additional features that distinguish animates versus artifacts. Some of these features may also be embedded in the parameters of the texture synthesis model. Multiple parameters (and their combinations) may create perceptual feature representations, including perceived curvature, that are useful for animate classification. Thus, this work opens up new avenues of research for isolating these different features, quantifying their relative contribution, and asking how they generalize to new image sets. Further, this approach is not limited to the model employed here: Other models of early visual processing (e.g., Balas, Nakano, & Rosenholtz, 2009) with similar parameters may also be useful to parametrically manipulate images and determine the role of other features in distinguishing between objects of different categories.

Another approach that is used to identify feature differences between categories is to build machine-learning algorithms that classify images on the basis of *image* features. This approach is powerful because it is possible to use high-throughput methods to examine a wide range of candidate features (Pinto & Cox, 2012). However, this method also has limitations for understanding human perception: Machine-vision algorithms will capitalize on any differences which distinguish between training images, but these features are not necessarily those extracted by the human visual system. For example, whereas deep convolutional networks can now rival human performance in basic-level recognition (and, by extension, animate classification; i.e., Iandola et al., 2016; Krizhevsky, Sutskever, & Hinton, 2012), in some cases these models use image features that are far different from the perceptual features that the human visual system uses (Nguyen, Yosinski, & Clune, 2015).

We see these approaches as complementary, and they can be combined in order to identify the critical features which distinguish categories: High-throughput machine learning can be used to identify candidate features which distinguish between categories *at the image level*, and our texture-synthesis approach can be used to examine whether these features contribute to human categorization and perception.

## Are these the same features used to detect animates in natural scenes?

Other studies have found a general advantage to detect animals versus inanimate objects in natural scenes (Crouzet, Joubert, Thorpe, & Fabre-Thorpe, 2012; New, Cosmides, & Tooby, 2007; Thorpe et al., 1996). Are the mid-level perceptual features we extracted with the texture synthesis model the same features that people use to rapidly find animals in natural scenes? In the present studies, we investigated the ability to detect an isolated picture of an animal among other isolated pictures of animals or objects. Thus, our results are not directly comparable to those from paradigms that examine the detection of animals in natural scenes (e.g., New et al., 2007).

Nonetheless, these mid-level features could explain part of how the visual system detects animates so rapidly. Previous work has found that the detection of animates in natural scenes is significantly impaired—though still rapid—when diagnostic parts (e.g., limbs, eyes, mouths) are eliminated (Delorme, Richard, & Fabre-Thorpe, 2010). Here, our texform stimuli eliminate these diagnostic features, while retaining differences in texture and form between animates and inanimates. One idea is that both diagnostic parts and mid-level features may play complementary roles in the detection of animals in natural scenes.

In addition, we found that mid-level texforms were classifiable as animates versus inanimates to different degrees— some were reliably classified as animates versus inanimates, whereas others were not. Thus, the speed with which animates are detected in natural scenes could increase as their mid-level features are more accurately classified as animate versus inanimate.

## When are perceptual differences important for conceptual processing?

The present findings suggest that mid-level perceptual representations contain early cues to an object's animacy, and previous work has drawn the same conclusion for objects that differ in real-world size (Long et al., 2016). Why might the adult visual system be sensitive to these mid-level cues? One idea is that these mid-level cues bypass basic-level recognition, allowing conceptual properties about an object to be inferred prior to basic-level recognition (i.e., Alaoui-Socé, Long, & Alvarez, 2015). For example, if the visual system can identify something as animate early in processing, the conceptual system can then infer that this object is likely to move on its own and have internal goals. Similarly, if the visual system can identify something as a small object, this is likely to be an object that can be picked up and used as a tool.

Consistent with this possibility, emerging evidence suggests that mid-level features feed-forward to influence high-level categorization in the absence of basic-level recognition. If task-irrelevant novel objects (i.e., greebles) are symmetric and have a typical animate skin color, adults categorize person attributes faster than object attributes (Cheung & Gauthier, 2014). In addition, if irrelevant texforms have animate features, adults categorize animal words faster than object words (Alaoui-Socé et al., 2015). Similarly, mid-level features trigger the processing of familiar, real-world size in the familiar-size Stroop task (Long & Konkle, under review).Thus, information about an object's animacy or real-world size, fed forward from mid-level feature analysis, may constrain the space of basic-level representations considered by the visual system during recognition. However, given that these featural cues are probabilistic (versus deterministic), they could sometimes lead the visual system astray. For example, the reason that some artifacts may look more like animals at first glance (e.g., a garden hose) could be because they have the mid-level features that lead the system to initially misclassify it as an animal (e.g., a snake). Indeed, in our image set, there were several texforms generated from artifacts that were routinely misclassified as animals.

Might other conceptual dimensions show these same effects? One longstanding idea is there may be privileged relationships between certain types of visual features and specific domains of conceptual processing (Caramazza & Shelton, 1998; Mahon & Caramazza, 2009). This account predicts that there may be privileged relationships between perceptual features and evolutionary salient distinctions—including the distinction between animates and artifacts. Accordingly, other work has found that object elongation is a reliable cue to whether something is a tool (Almeida et al., 2014). Importantly, this account does not necessitate that these feature representations are innate; rather, these conceptual dimensions could become mapped to different features early in development.

This account also predicts that mid-level features will *not* bypass basic-level recognition if they are cues to a relatively arbitrary distinction. Arbitrary distinctions may only cover a limited slice of possible object categories and may not be relevant for how we interact with objects. For example, handmade versus manufactured objects could have different mid-level features. However, as we don't interact with all handmade objects in reliably different ways than manufactured objects, there may be no incentive for these (hypothetical) mid-level cues to feed-forward to the semantic system. On the other hand, though some conceptual dimensions could be highly relevant for how we interact with objects (e.g., caloric value) they may not have perceptual correlates. Ultimately, however, these questions are best left to future research, and the present study provides an approach and framework for testing these hypotheses.

## Conclusion

Across a series of experiments, participants searched faster and more efficiently for a target when distractors differed in animacy. This was true both when stimuli were highly controlled and when reduced to unrecognizable, mid-level texforms. Moreover, electrophysiological recordings revealed that this difference was present at the early stage of target individuation. Thus, animates versus inanimates have different mid-level perceptual features that may be extracted early during the process of object recognition. Broadly, examining if different conceptual dimensions have mid-level cues may deepen our understanding of how the visual system connects early perceptual inputs with semantic representations.

*Keywords: object recognition, animacy, mid-level perceptual features, EEG/ERP, N2pc*

## Acknowledgments

Bria Long, Viola S. Störmer, and George A. Alvarez designed the study concept. Bria Long and Viola S. Störmer conducted data collection and analyses. Bria Long drafted the manuscript, and all other authors provided revisions and approved the final draft of the manuscript. J. Freeman provided the code that was used to generate the texform stimuli. This work was supported by an NSF CAREER grant (BCS-0953730) to G. A. A. and a Marie-Curie fellowship to V. S. S. (EU Grant PIOF-GA-2012-329920).

Commercial relationships: none.

Corresponding author: Bria Long.
Email: brialorelle@gmail.com.
Address: Department of Psychology, Harvard University, Cambridge, MA, USA.

## Footnotes

[1] Though one texform of a ship was recognizable 53% of the time, basic-level recognition accuracy for all other texforms was 20% or below.

[2] In addition, there were no differences in voltage when both the targets and distractors were animal texforms (uniform displays, animal targets) versus when both were manmade object texforms (uniform displays, object targets): uniform animals, $M = 1.85\ \mu V$; uniform objects, $M = 2.18\ \mu V$, $t(15) = -1.07$, $p = 0.303$.

[3] As participants were more accurate on mixed-animacy trials, there were more trials available for analysis in the mixed-animacy relative to uniform-animacy conditions. To confirm that this imbalance in the number of trials did not spuriously create this result, we matched the number of mixed and uniform trials in each subject that contributed to the n2pc analyses, and found the same pattern of results (interaction between display type and n2pc, $F(1, 15) = 6.69$, $p = 0.02$).

[4] We thank an anonymous reviewer for this suggestion.

## References

Alaoui-Socé, A., Long, B., & Alvarez, G. A. (2015). Animate shape features influence high-level animate categorization. *Journal of Vision*, *15*(12): 1159, doi:10.1167/15.12.1159. [Abstract]

Almeida, J., Mahon, B., Zapater-Raberov, V., Dziuba, A., Cabaço, T., Marques, J. F., & Caramazza, A. (2014). Grasping with the eyes: The role of elongation in visual recognition of manipulable objects. *Cognitive, Affective and Behavioral Neuroscience*, *14*(1), 319–335.

Antal, A., Keri, S., Kovacs, G., Janka, Z., & Benedek, G. (2000). Early and late components of visual categorizations: an event-related potential study. *Cognitive Brain Research*, *9*(1), 117–119.

Balas, B. J. (2006). Texture synthesis and perception: Using computational models to study texture representations in the human visual system. *Vision Research*, *46*(3), 299–309.

Balas, B. J. (2012). Contrast negation and texture synthesis differentially disrupt natural texture appearance. *Frontiers in Psychology*, *3*, 515.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12)13, 1–18, doi:10.1167/9.12.13. [PubMed] [Article]

Banno, H., & Saiki, J. (2015). The use of higher-order statistics in rapid object categorization in natural scenes. *Journal of Vision*, *15*(2):4, 1–20, doi:10.1167/15.2.4. [PubMed] [Article]

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition*, *14*(2), 201–212.

Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 20–30.

Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, *129*, 255–263.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34.

Cheung, O. S., & Gauthier, I. (2014). Visual appearance interacts with conceptual knowledge in object recognition. *Frontiers in Psychology*, *5*, 793.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., Abdi, H. & Haxby, J. V. (2012). The representation of biological classes in the human brain. *The Journal of Neuroscience*, *32*(8), 2608–2618.

Crouzet, S. M., Joubert, O. R., Thorpe, S. J., & Fabre-Thorpe, M. (2012). Animal detection precedes access to scene category. *PLoS One*, *7*(12), e51471.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317–329.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in Psychology*, *1*, 21.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Science*, *11*(8), 333–341.

Dowdall, J. R., Luczak, A., & Tata, M. S. (2012). Temporal variability of the N2pc during efficient and inefficient visual search. *Neuropsychologia*, *50*(10), 2442–2453.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201.

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*(7), 974–981.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123–144.

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). *Digital image processing using MATLAB*. S1: Gatesmark Publishing.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition as soon as you know it is there, you know what it is. *Psychological Science*, *16*(2), 152–160.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210.

Hwang, A. D., Wang, H. C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, *51*(10), 1192–1205.

Iandola, F. N., Moskewicz, M. W., Ashraf, K., Han, S., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <1MB model size. arXiv preprint arXiv, 1602.07360.

Johannes, S., Münte, T. F., Heinze, H. J., & Mangun, G. R. (1995). Luminance and spatial attention effects on early visual processing. *Cognitive Brain Research*, *2*(3), 189–205.

Julian, J. B., Ryan, J., & Epstein, R. A. (2016). Coding of object size and object category in human visual cortex. *Cerebral Cortex*, bhw150, e-pub ahead of print.

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: Event-related potentials during picture and word categorization. *Memory & Cognition*, *29*(1), 100–116.

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *The Journal of Neuroscience*, *33*(25), 10235–10242.

Konkle, T., & Oliva, A. (2012a). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114–1124.

Konkle, T., & Oliva, A. (2012b). A familiar-size Stroop effect: Real-world size is an automatic property of object representation. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 561–569.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105). Cambridge, MA: MIT Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Levin, D. T., Takarae, Y., Miner, A. G., & 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*(4), 676–697.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 565–572.

Long, B., & Konkle, T. (Under review). A familiar-size Stroop effect in the absence of basic-level recognition. Under review.

Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. *Journal of Experimental Psychology: General*, *145*(1), 95–109.

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 1000–1014.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 1057.

Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research*, *181*(3), 531–536.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason*, *4*(2), 61–64.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*(2), 507–517.

Naber, M., Hilger, M., & Einhäuser, W. (2012). Animal detection and identification in natural scenes: Image statistics and emotional valence. *Journal of Vision*, *12*(1):25, 1–24, doi:10.1167/12.1.25. [PubMed] [Article]

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences, USA*, *104*(42), 16598–16603.

Nguyen, A., Yosinski, J., & Clune, J. (2015, June). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Computer Vision and Pattern Recognition* (2015 IEEE Conference, pp. 427–436). New York: IEEE.

Okazawa, G., Tajima, S., & Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proceedings of the National Academy of Sciences*, *112*(4), E351–E360.

Pinto, N., Cox, D.D. (2012) High-throughput-derived biologically-inspired features for unconstrained face recognition. *Image and Vision Computing*, *30*(3), 159–168.

Proverbio, A. M., Del Zotto, M., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience*, *8*(1), 24.

Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, *459*(3), 142–146.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283.

Störmer, V. S., Winther, G. N., Li, S. C., & Andersen, S. K. (2013). Sustained multifocal attentional enhancement of stimulus processing in early visual areas predicts tracking performance. *Journal of Neuroscience*, *33*(12), 5346–5351.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition*, *17*(1), 159–170.

Võ, M. L. H., & Wolfe, J. M. (2013). Differential ERP signatures elicited by semantic and syntactic processing in scenes. *Psychological Science*, *24*(9), 1816.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684.

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 121–138.

Wu, C. C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*, 54.

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision*, *14*(1):28, 1–20, doi:10.1167/14.1.28. [PubMed] [Article]