Visual anagrams reveal high-level effects with ‘identical’ stimuli

Tal Boger, Chaz Firestone
Department of Psychological and Brain Sciences, Johns Hopkins University
tboger1@jhu.edu, chaz@jhu.edu

A fundamental question in psychology and neuroscience concerns how the mind represents not only lower-level stimulus features, such as luminance, contrast, or spatial frequency, but also richer, higher-level properties, such as animacy, emotion, or real-world size. Numerous findings suggest that such high-level properties are encoded automatically[1,2], engage visual attention[3,4], and organize neural responses[5,6]. However, a critical challenge arises when interpreting such findings: High-level categories systematically covary with lower-level features, such that effects attributed to high-level properties may instead be driven by their lower-level covariates. Can this challenge be overcome? Here, we introduce a novel approach by leveraging ‘visual anagrams’ — a diffusion-based technique for generating images whose interpretations change radically with orientation; for example, a cow when upright and a mouse when inverted[7]. Using real-world size as a case study, we generated anagrams depicting a canonically large object in one orientation and a canonically small object in another, and placed them in classic experimental paradigms. Five experiments revealed that many (but not all) effects of real-world size persisted under such exacting control. Together, our findings address a longstanding challenge in perception research and establish a broadly applicable tool for psychology and neuroscience.

Consider the rabbit and elephant in Figure 1A. Although they occupy roughly the same amount of space on the page, they differ in their real-world size. An extensive body of research suggests that this high-level difference is actively represented by the mind: Real-world size intrudes on orthogonal perceptual judgments[1,2], drives visual search[4], and constrains cortical representation[5]. But real-world size is not the only feature distinguishing the rabbit and elephant: They also differ in shape, curvature, spatial frequency, viewing angle, and other mid- and low-level properties. Thus, while differences in representation of these objects may arise from differences in real-world size, they may instead arise from correlated lower-level differences (an especially salient possibility given similar findings with distorted, unrecognizable stimuli[2,4,8]). Despite progress on this problem[6,9], isolating high-level properties from lower-level features remains an enduring challenge.

Now consider the rabbit and elephant in Figure 1B. These are actually the very same image, rotated 90°. They are ‘visual anagrams’, created using a diffusion-based technique that generates static images whose interpretations change radically when rotated[7]. The two images are pixel-wise identical, thus differing in a high-level property (here, real-world size) without differing in curvature, spatial frequency, luminance, contrast, and so on.

Version: 8/17/25; forthcoming, *Current Biology.*

Here, we exploit this technique to investigate high-level effects with otherwise 'identical' stimuli, eliminating nearly all lower-level covariation associated with conventional approaches. We generated images depicting a large object in one orientation and a small object in another (for example, rabbit-elephant, butterfly-bear), placed them in classic paradigms exploring real-world size (https://perceptionresearch.org/anagrams), and asked whether the original findings persist under these conditions.

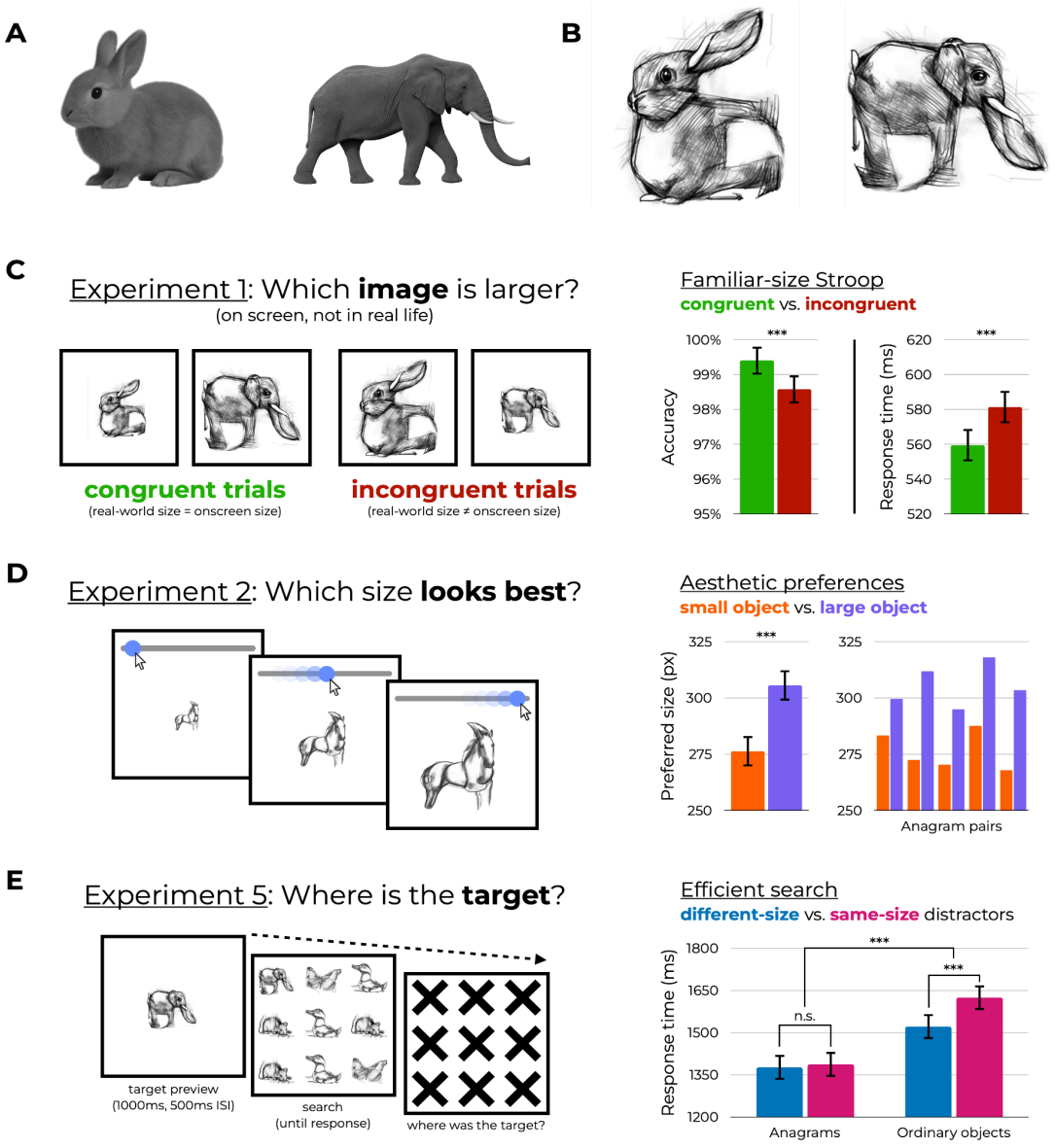

Figure 1. High-level effects with visual anagrams. (A) This rabbit and elephant differ in a high-level property — real-world size — but also in several mid-level and low-level properties, such as curvature, spatial frequency and contrast. (B) This rabbit and elephant are 'visual anagrams'[7]; they also differ in real-world size, but contain identical pixels (being the same image rotated 90°). (C) The familiar-size Stroop effect arose with visual anagrams (Experiment 1). (D) Real-world size drove aesthetic preferences with visual anagrams (Experiment 2). (E) Visual search was not facilitated by real-world size when using visual anagrams, although previously reported effects arose with non-anagram stimuli.

We first investigated automatic encoding of real-world size using the familiar-size Stroop task[1]. In this task, two images are displayed at different sizes, and subjects must say which is larger on the screen. Despite real-world size being explicitly task-irrelevant, performance is better when displayed size is congruent with real-world size (for example, rabbit-small, elephant-big). Experiment 1 adapted this design to our anagram stimuli. Consistent with previous work, we found a familiar-size Stroop effect (Figure 1C): Subjects were faster and more accurate on congruent trials than incongruent trials (–21.9 ms, $t(50) = 4.75$, $p < 0.001$; +0.8%, $t(50) = 3.80$, $p < 0.001$), even when the images were simply rotated versions of one another.

We next explored a connection between real-world size and aesthetic preferences. Previous work suggests that observers prefer canonically small objects to be displayed small, and canonically large objects to be displayed large[8,10]. Consistent with this work, Experiment 2 revealed that subjects preferred canonically large objects to be displayed larger than canonically small objects, even with visual anagrams (+29.3px, or +9.6%, $t(197) = 8.60$, $p < 0.001$; Figure 1D).

Whereas Experiments 1 and 2 included a familiarization phase in which subjects first matched category labels to the anagram stimuli, Experiments 3 and 4 replicated those experiments without this phase. The same patterns emerged (Stroop: –31.7 ms, $t(45) = 5.58$, $p < 0.001$; Preferred size: +24.3px, or +8.2%, $t(197) = 8.23$, $p < 0.001$), replicating our results and demonstrating that visual anagrams are readily identifiable without prompting.

Finally, we investigated links between real-world size and attention. Previous work reports that targets are easier to locate when their real-world size differs from distractors[4]. Using that paradigm, however, Experiment 5 found little-to-no effect with anagram stimuli (11.1 ms advantage, $t(48)=0.51$, $p = 0.61$, BF10 = 0.176; Figure 1E), suggesting that the original findings may indeed be driven by correlated lower-level properties. Importantly, Experiment 5's design replicated earlier search findings using non-anagram stimuli[4]; those stimuli successfully reproduced previously reported effects (102.6 ms advantage, $t(48) = 4.89$, $p < 0.001$), which were significantly stronger than the (non-significant) effects with anagrams (91.5ms difference, $t(48) = 3.68$, $p < 0.001$).

Our work confronts the longstanding challenge of disentangling high-level properties from lower-level covariates. Our results suggest that real-world size per se is represented by the mind: It is encoded automatically and drives aesthetic judgments, in ways that go beyond its lower-level correlates. Not all effects persisted in this way, however, highlighting how this approach can both support and reframe high-level psychophysical effects.

These findings build on previous work showing that many real-world size effects occur even with unrecognizable 'texforms' that preserve mid-level features such as curvature[2,4,8]. That work raises the question of whether real-world size effects are fully captured by such features or

instead go beyond them. Experiments 1–4 suggest that there are indeed effects that go beyond mid-level stimulus features, whereas Experiment 5 suggests that at least some effects are driven mostly or only by such features (in ways that are nevertheless consistent with the original claims).

Importantly, our approach is perfectly general. Though we manipulated real-world size, one could generate anagrams of happy faces and sad faces, tools and non-tools, or animate and inanimate objects, overcoming low-level confounds associated with such stimuli[3,6]. The present work thus serves as a 'case study', yielding concrete discoveries about real-world size and validating a broadly applicable tool for psychology and neuroscience.

References

1. Konkle, T., and Oliva, A. (2012). A familiar-size Stroop effect: Real-world size is an automatic property of object representation. J. Exp. Psychol. Hum. Percept. Perform. *38*, 561-569.

2. Long, B., and Konkle, T. (2017). A familiar-size Stroop effect in the absence of basic-level recognition. Cognition *168*, 234-242.

3. Levin, D. T., Takarae, Y., Miner, A. G., and Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. Percept. Psychophys. *63*, 676-697.

4. Long, B., Konkle, T., Cohen, M. A., and Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. J. Exp. Psychol. Gen. *145*, 95-109.

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

6. Proklova, D., Kaiser, D., and Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. J. Cogn. Neurosci. *28*, 680-692.

7. Geng, D., Park, I., and 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.

8. Chen, Y. C., Deza, A., and Konkle, T. (2022). How big should this object be? Perceptual influences on viewing-size preferences. Cognition *225*.

9. Schmidt, F., Kleis, J., Morgenstern, Y., and Fleming, R. W. (2020). The role of semantics in the perceptual organization of shape. Sci. Rep. *10*.

10. Konkle, T., and Oliva, A. (2011). Canonical visual size for real-world objects. J. Exp. Psychol. Hum. Percept. Perform. *37,* 23-37.

Supplemental Information: Visual anagrams reveal high-level effects with 'identical' stimuli

Tal Boger, Chaz Firestone

This document further explicates the methods of the experiments described in the main text. All of our pre-registrations, stimuli, experiment scripts, anonymized data, and analysis code are available on OSF (https://osf.io/trsqj/). Readers can experience all the tasks for themselves at https://perceptionresearch.org/anagrams.

Supplemental methods (all experiments)

All experimental designs, sample sizes, exclusion criteria, and analysis plans were pre-registered. Subjects were recruited via the online platform Prolific (https://www.prolific.co/). Unique participants were recruited for each experiment.

Stimulus generation

Images were generated using the Advanced Research Computing at Hopkins core facility (ARCH), with the "visual anagrams" model[S1]. (Though the model is named "visual anagrams", readers might recognize these images as ambigrams, following Douglas Hofstadter's coinage[S2].) We created a list of many large and small object labels, and then prompted the model to produce many combinations of every possible pair under different degrees of rotation (clockwise 90°, counterclockwise 90°, and 180°). The model allows the two desired interpretations to be specified in advance, and the best outputs it produces are recognizable without special prompting. (These aspects of visual anagrams, combined with the ability to automate their generation, allow them to go beyond extant approaches for separating high- and low-level features—including special bistable figures such as a duck-rabbit or face-vase, contextual influences on object recognition, or attentional modulation of 3D orientation.)

From these images, we selected (before running any of our experiments) anagrams that seemed to be particularly identifiable as both prompts, and used them in all of our experiments. This resulted in 10 images, or 5 anagram pairs: rabbit-elephant, butterfly-bear, duck-horse, mouse-cow, and lighter-truck. These stimuli are available in our OSF repository. Each member of an anagram pair is pixelwise-identical subject to rotation.

Of course, even containing the exact same pixels does not (and could not) hold every mid- and low-level feature constant. For example, rotating the anagrams changes the orientation of edges in the stimulus, and can also change an image's aspect-ratio. It would, of course, be impossible to hold all mid- and low-level features constant without using the exact same image (without

transformation); but in that case, there would be no effects to discover (barring contextual or attentional effects). In other words, while this approach does not allow complete control of mid- and low-level features, it allows for more control than has previously been possible while systematically altering the high-level property of interest.

Supplemental methods for Experiment 1: Familiar-size Stroop

This experiment used the familiar-size Stroop task[S3] with visual anagrams.

Participants

50 participants were recruited (though 51 participants ended up completing the task due to a feature of Prolific's recruiting process).

Procedure

On each trial of the experiment, subjects were shown two images side by side; one was displayed larger than the other, and subjects simply had to say which one was larger on the screen. The images always included one randomly selected canonically small object and one randomly selected canonically large object (both from the anagrams stimulus set we generated); but participants were instructed to judge only the display size of the images. Importantly, on half of trials, the visual size of the images was congruent to their real-world size (i.e., the real-world-larger object was larger on the screen); and on the other half of trials, visual size was incongruent to real-world size (i.e., the real-world-smaller object was larger on the screen).

The experiment contained two blocks. In one, participants judged which of two images was larger; in the other, they judged which of two images was smaller. The order of the two blocks was randomized for each participant. Within each block, there were 100 test trials, which was reached through all combinations of 5 small objects × 5 large objects × 2 congruence conditions × 2 locations (larger image on the left or larger image on the right). Each block also contained 8 'catch' trials which depicted circles and triangles of clearly differing sizes; 4 of these trials were placed at the start of each block, and the remaining 4 were randomly interspersed throughout. Subjects responded using their keyboard as to which image (left or right) was larger on the screen. The visually larger image was always 250px in height and width, and the smaller image was always 150px in height and width.

This experiment (and Experiment 2) contained an additional page at the start of the experiment which asked subjects to match the anagram images to their corresponding labels. On this page, subjects were shown all 10 anagram images at once, and then were given object labels

one-by-one and asked to click on the image matching each label (e.g., “click on the rabbit”, “click on the horse”).

As per our pre-registration, we excluded subjects who failed to submit a complete dataset or who did not respond correctly on at least 75% of catch trials. This excluded 0 participants. Next, we excluded trials with a response time below 200ms or above 1500ms (as in previous work[S2]). This excluded 2.7% of trials (301/11016).

Supplementary Analyses

Beyond the analyses reported in the main text (which collapse over all trials), we also analyzed the subset of trials in which the two images shown to subjects were anagrams of one another. This was pre-registered as a secondary analysis; it revealed better performance for congruent trials than incongruent trials (–25.7ms, $t(50)=3.42$, $p<.01$).

Supplemental methods for Experiment 2: Aesthetic preferences

This experiment tested the effect of real-world size on aesthetic preferences (specifically, viewing size, as in previous work[S4]) in visual anagrams.

Participants

200 subjects were recruited.

Procedure

Subjects were instructed to adjust the size of an image to be their preferred size for viewing it. On each trial, subjects freely moved a slider which changed an image’s size. Once the image reached the size at which they thought it looked best, subjects could complete the trial and proceed. The slider was always initialized in the center; and in its initial state, no image appeared (i.e., the image only appeared and began to change size once the slider was adjusted). The slider could be used to adjust the image’s size from 0px to 400px.

Subjects completed 10 trials: 1 for each individual anagram object. In other words, subjects adjusted the size of 5 canonically small objects, and their 5 canonically large anagram counterparts. The order of the trials was randomized for each subject. As in Experiment 1, this experiment contained a labeling page at the start of the experiment.

As per our pre-registered analysis plans, we excluded 2 subjects who failed to submit a full data set.

Supplemental methods for Experiments 3 and 4: No labeling

Experiments 3 and 4 were direct replications of Experiments 1 and 2. The only difference in these experiments is that there was no labeling phase at the start of the experiment—otherwise, Experiment 3 was exactly identical to Experiment 1, and Experiment 4 was exactly identical to Experiment 2 (not only in terms of design, but also in terms of exclusion criteria and analysis). The goal of these experiments was to ensure that (a) the anagrams are readily identifiable as the intended objects without further prompting, and (b) the labeling phase was not biasing subjects in any way.

Supplemental methods for Experiment 5: Efficient visual search

Experiment 5 asked whether real-world size drives efficient visual search—even with anagrams—using the design and timing parameters from Long et al. (2016)[S5].

We ran Experiment 5 after piloting three earlier variations of this study. Two of these three variations failed to yield significant effects with visual anagrams; however, one variation produced a marginally significant effect ($.01 < p < .05$) that was several times smaller than with non-anagram stimuli. This led us to pre-register and run Experiment 5, to directly compare effects with anagram and non-anagram stimuli; as we report in the main text, this experiment again failed to yield significant effects with visual anagrams. Overall, it seems possible that there are effects of real-world size on visual search when lower-level features are controlled; however, if they do exist they are likely weaker than the original effects—suggesting that the original effects are explained in whole or in part by such correlated features (whether mid-level features such as curvature, texture, and shape, or low-level features such as luminance, contrast, and spatial frequency). Note that this conclusion needn't be inconsistent with Long et al.'s (2016) own, and indeed may be complementary. Long et al. claim that large and small objects are distinguished by mid-level features, and demonstrate this by showing facilitated search for texforms derived from small objects appearing among texforms derived from large objects (and vice versa). This finding demonstrates that mid-level features distinguish large and small objects, but leaves open the question of whether other information does so as well; our studies essentially ask whether there are any differences 'left over' once most or all such features are controlled.

Participants

50 subjects were recruited.

Procedure

Subjects completed a visual search task whose timing and task procedure followed Long et al. (2016). A target object was presented for 1000ms, after which it disappeared and there was a 500ms inter-stimulus interval. Then, the target object appeared among 8 distractors randomly placed in an invisible 3-by-4 grid. Subjects were instructed to press the spacebar as soon as they located the target; after this keypress, all the images were replaced by large "X"s. Subjects then used their mouse to click on the "X" that occupied the location of the target.

Importantly, in one half of trials, the distractors were all of a similar real-world size as the target (e.g., all were small or all were large). In the other half of trials, the distractors' real-world size differed from that of the target (e.g., the target was real-world-small and the distractors were real-world-large, or vice versa).

The experiment consisted of 2 blocks. One block contained the visual anagrams. The other contained natural images of real objects drawn from Konkle and Oliva (2012). Following Long et al. (2016), we matched these images for luminance, spatial frequency, etc.[S5]. The order of the two blocks was randomized for each subject.

Each block consisted of 100 test trials, which were equally split among different-size distractor and same-size distractor trials (50 each). The 50 trials of each type were themselves equally split between the target being a real-world small object and the target being a real-world large object. Distractors were randomly chosen, with one key rule: On anagram trials, the distractor could not be an anagram of the target (e.g., if the target is the rabbit, then the elephant cannot be among the distractors). Each block also began with 4 'catch' trials in which the targets and distractors were simple triangles and circles. There were 4 additional catch trials randomly interspersed throughout each block.

As per our pre-registered analysis plans, we excluded subjects who did not respond correctly on at least 75% of catch trials. This excluded 1 subject. Then, we excluded trials with a response time below 200ms or above 5000ms, which excluded 4.1% of trials (436/10584).

Supplementary Analyses

In addition to the (pre-reregistered) analyses reported in the main text, we also conducted a Bayes factor analysis to evaluate the strength of evidence that there is no effect with visual anagrams. This analysis was exploratory. It revealed moderate-to-strong evidence in favor of the null hypothesis (i.e., no search advantage for anagrams; BF10=0.176). We thank a reviewer for comments that led to this analysis.

Additionally, we make the following observation: Search performance is *faster* for the anagram stimuli (mean RT=1385ms) than non-anagram stimuli (mean RT=1573; $t(48)=4.09$, $p<.001$).

This suggests that the anagram stimuli were not particularly difficult to search for, which was perhaps a potential worry given their sketch-like characteristics. Again we thank a reviewer for prompting this follow-up analysis.

Author contributions

T.B. and C.F. contributed equally to the study's conceptualization, methodology, visualization, writing—original draft, and writing—review and editing. T.B. served as lead for data curation and formal analysis. C.F. served as lead for project administration and supervision.

Supplemental References

S1. Geng, D., Park, I., and 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.

S2. Hofstadter, D. (2025). Ambigrammia: Between Creation and Discovery (Yale University Press).

S3. Konkle, T., and Oliva, A. (2012). A familiar-size Stroop effect: real-world size is an automatic property of object representation. J. Exp. Psychol. Hum. Percept. Perform. *38*, 561-569.

S4. Konkle, T., and Oliva, A. (2011). Canonical visual size for real-world objects. J. Exp. Psychol. Hum. Percept. Perform *37,* 23-37.

S5. Long, B., Konkle, T., Cohen, M. A., and Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. J. Exp. Psychol. Gen. *145*, 95-109.