

A variational image reconstruction algorithm reveals distortion and uncertainty in mental imagery

Thomas Naselaris (tnaselar@musc.edu)

Medical University of South Carolina, 96 Jonathan Lucas Street
Charleston, SC 29425, USA

Abstract

Subjective reports on mental images may not be consistent with the appearance of seen images (distortion), and may not be consistent with each other (uncertainty). We introduce a method for estimating the distortions and uncertainty associated with mental images from subjective reports. Application of the method to a small preliminary dataset suggests that distortion and uncertainty in mental imagery arise from overscaling and undersampling imagined objects.

Keywords: mental imagery; variational inference

Summary

How complex can a mental image be? Introspection suggests a limit. Imagine a white candle on a tabletop in an otherwise empty room (MacKisack et al., 2016). Now imagine two candles, eight candles, now sixteen, each a different size. In each iteration of the mental image, identify the smallest candle. Determine if it is more than half the size of the largest candle. Determine its location. These simple determinations become increasingly difficult to make as the mental image becomes more complex. For some level of complexity (twenty candles, a hundred) the determinations become guesses. These guesses may or may not be consistent across repeated inspections of a mental images. If consistent, the guesses might simply reflect internal biases that distort the way that objects appear in the real world. If inconsistent, the guesses would indicate objective uncertainty about what we think we are imagining.

Here we describe a method for quantitatively characterizing the uncertainty and distortions associated with complex mental images. We call the method `viral` as it is a variational image reconstruction algorithm. `viral` can in principle be applied to brain activity measured during mental imagery; however, any measure of the uncertainty and distortions associated with mental images is likely to become conflated with our own distorted and uncertain understanding of what neural populations represent. Therefore, we apply the method here to multiple, independently sampled behavioral reports about the local content of subjects' mental images (Figure 1, (Podgorny & Shepard, 1978)). From these subjective reports `viral` infers a posterior distribution over pixelwise segmentations of the subjects' mental images (Fig. 3). We show that inferred segmentations can be used to accurately predict subjective reports. We also use them to measure the size of imagined objects relative to seen ones. Preliminary data indicate that in simple mental images, imagined object size is nearly

identical to seen object size. However, in complex mental images the imagined foreground objects are much larger than seen ones (Fig. 5). In addition to the variational posterior `viral` infers two independent measures of uncertainty: a hallucination rate and a forgetting rate (Figure 2). The hallucination rate quantifies how likely the subjects are to hallucinate objects that were absent in previous reports, while the forgetting rate quantifies how likely subjects are to forget objects that were present. Our preliminary data indicate that the forgetting rate increases abruptly with the complexity of mental images, while the hallucination rate is negligible for even the most complex mental images (Fig. 6). These preliminary findings suggest that distortions and uncertainty associated with mental imagery arise from subjects overscaling and undersampling the foreground objects near the inspected region of their mental images. We speculate that these forms of distortion and uncertainty are related to the low spatial resolution of the high-level visual brain areas that support mental imagery (Breedlove, Naselaris, & St-Yves, 2017).

Methods

Experiment

We present a preliminary analysis of data from three subjects. Each subject participated in a "paint-by-numbers" experiment (Figure 1, top) that required them to give multiple subjective reports on the contents of two to four seen or imagined target images (Figure 1, bottom). Each subjective report was a count of the number of objects that underlay a polygonal probe presented on a blank screen during imagery experiments or overlaid onto the target image during vision experiments. See Figure 1 for more detail.

Analysis

Data for each subject, target image and type of experiment (vision or imagery) were analyzed separately. For each dataset we inferred a variational posterior $q(Z)$ over images Z and two uncertainty parameters θ_+, θ_- . See Figure 2 for more detail. The variational posterior allowed us to measure and compare the sizes of seen and imagined objects for same subject and target image. In this preliminary study we present a representative set of examples (Fig. 5).

Results

Validation of the method

The subjective reports gathered during the experiment are local object counts. Object count is an inherently ambiguous statistic, as it does not specify what objects are locally present,

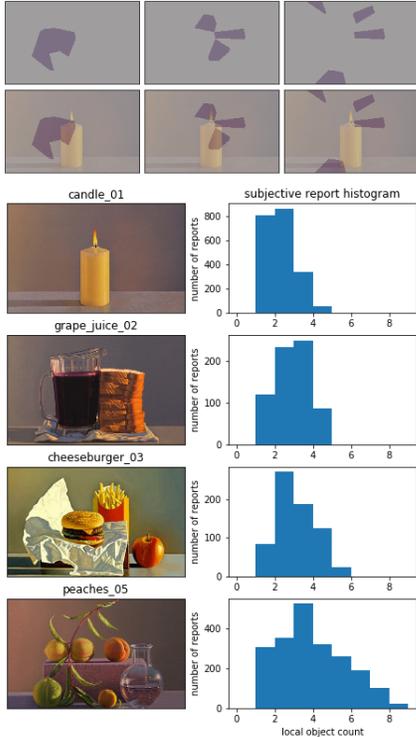


Figure 1: A paint-by-numbers experiment. (Top) What the subjects saw during the experiment. Subjects are given several minutes to become familiar with a painting (in this example “Still life with candle” by David Ligare) presented on a computer monitor. The painting disappears, but subjects are instructed to imagine that it is still present onscreen. Subjects are then presented with a series of contiguous and non-contiguous semi-transparent polygonal probes and must count and report with a keypress the number of *imagined* objects that underlay each probe. 345 object counts are obtained per image per subject. The painting then reappears and remains onscreen and the probes are again displayed (same probes, different order). The subjects repeat the experiment, this time counting the number of *seen* objects that underlay each probe. Subjects are instructed to count the background as a distinct object; no other guidance is provided about what parts of the target image are to be treated as distinct objects. (Bottom left) The paintings that the subjects imagined and inspected (see Acknowledgements for art credits). The paintings vary in complexity. Tags indicate prominent foreground object and ordinal complexity (e.g., candle-01, peaches-05). Two of the three subjects imagined and inspected only the least (candle-01) and the most (peaches-05) complex paintings. A third subject imagined and inspected all four. (Bottom right) Subjective report data. Each histogram tallies the object counts reported for the indicated painting (combined data from all subjects, imagery and vision experiments)

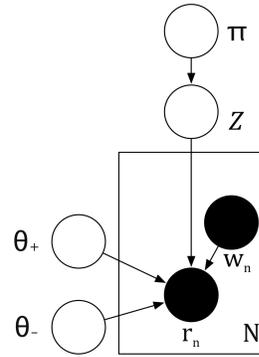


Figure 2: The generative story. The probabilistic graphical models depicts our assumptions about how visual or mental images Z generate subjective reports r_n in response to a polygonal probe w_n . In *viral*, the image is simply an assignment of a discrete object label to each of the pixels in Z . The variable $\pi = (\pi_1, \dots, \pi_K)$ parameterizes the prior assignment probabilities for each of the K objects in Z . Whenever a probe w_n is observed by the subject, the subject surveys the K objects in Z . If an object underlays the probe, the subject ignores it with probability $1 - \theta_+$, which is the forgetting rate. If an object does not underlay the probe, the subject counts it anyway with probability θ_- , which is the hallucination rate. The plate indicates that each of the $N = 345$ subjective reports obtained for a visual or mental image are treated as independent of each other given Z . *viral* is simply coordinate ascent variational inference (Blei et al., 2016) applied to this graph. Thus *viral* yields an approximate (variational) posterior over visual or mental images $q(Z) \approx p(Z|r_1, \dots, r_N, w_1, \dots, w_N, \theta_+, \theta_-)$. We assume that $q(Z)$ factors across pixels of Z . We obtain point estimates of the forgetting and hallucination rates by maximizing the variational posterior with respect to θ_+, θ_- .

nor how the pixels that depict those objects are distributed over space. It is therefore important to validate that *viral* can recover meaningful information about the location and size of seen or imagined objects in a target image. Simulations revealed that for sufficiently large numbers of low-noise subjective reports ($n \geq 1,000$, see Fig. 3, second panel from left) *viral* learns an approximate posterior distribution $q(Z)$ over images Z that very accurately delimits the objects in the target image. That is, if we identify all the pixels in a target image that belong to a single object, there will be some object label k such that $q(z_d = k) \approx 1$ whenever the pixel z_d belongs to that object, and $q(z_j = k) \approx 0$ whenever the pixel z_j does not. Of course, experiments with human subjects yield fewer, noisier datapoints. The distributions corresponding to the “candle” object for one subject during vision (Fig. 3, third from left) and mental imagery (Fig. 3, fourth from left) appear accurate. For more complex images, however, the correspondence between distributions and the target images is harder to appreciate by

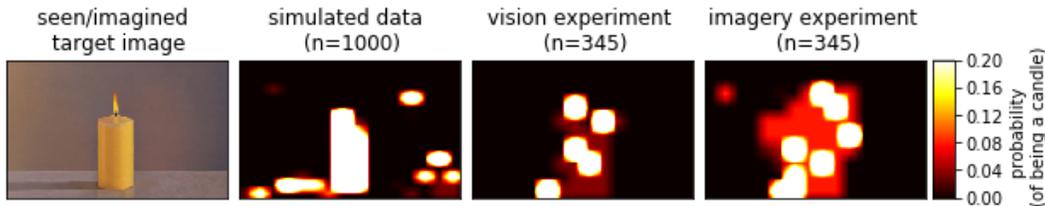


Figure 3: Reconstructing seen and imagined candles from subjective reports. At left is the target image that the subjects either saw or imagined during the experiment. Given their subjective reports, `viral` yields a variational posterior $q(Z)$ over pixelwise segmentations of the image. This means that for each pixel z_d in Z , the variational posterior probability that z_d belongs to k^{th} object is $q(z_d = k)$. The three panels to the right of the target image show the probability that each pixel is part of the candle. Given simulated subjective reports (second from left; hallucination rate = 0.005; forgetting rate = 0.01) the candle is recovered almost perfectly (the false positives in the periphery result from the inherent ambiguity of the subjective reports). Given reports from a real subject observing the target image (third from left) or imagining (fourth from left) the spatial distribution still clearly resembles the candle, although it is more diffuse for the imagined candle.

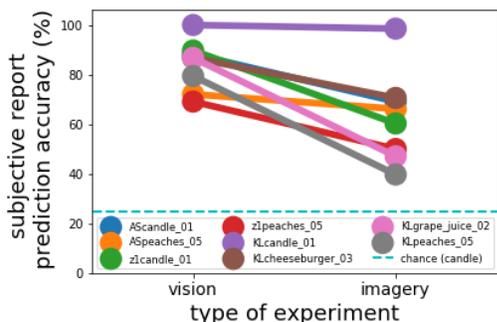


Figure 4: Validation of `viral`. Data for each subject was split into training and testing subsets containing 80% and 20% of the data, resp. We validated `viral` for each subject, visual and mental image by using it to predict subjective reports (i.e., discrete object counts) on the test subset. Let $q_{\text{train}}(Z) \approx p(Z|r_1, \dots, r_N, w_1, \dots, w_N, \theta_+^*, \theta_-^*)$ be the variational posterior inferred by observing the training data only, with θ_+^*, θ_-^* our point estimate of the uncertainty parameters. The predictive distribution for subjective report r_i in response to probe w_i in the testing subset is $p(r_i|w_i) = \mathbb{E}_{q_{\text{train}}(Z)}[p(r_i|w_i, Z)]$. Above, we show the percentage of time that the argmax of this predictive distribution matches the subjective reports in the testing subset. Lines connect the results obtained for the visual and imagery experiments for the same subject and target image. The dashed line indicates chance for the candle target image (an upper bound on chance accuracy). For all subjects and targets `viral` yields an accurate cross-validated predictive distribution. Perhaps not surprisingly, prediction accuracy is worse for mental imagery than for vision for most subjects and target images, and is worst of all for complex mental images (e.g., peaches-05).

eye. Thus, we used the variational posterior for each target image to derive a predictive distribution $p(r_n|w_n)$, where r_n

is the subjective report, and w_n is the probe presented on the n^{th} trial. We found that even for complex images (such as peaches-05) the mode of the predictive distribution gave an accurate (i.e., greater than chance) prediction of subjective reports (Fig. 4). These results endow `viral` with face- and empirical validity, and licence us to use it as a tool for investigating subjects' representations of both seen and mental images.

Distortions of imagined objects

How do seen and imagined objects diverge as the target image becomes more complex? Here, we investigate a potential divergence in the size of seen and imagined objects. For a given object in a target image—call it object k —we calculate the probability under $q(Z)$ that the seen or mental image contains an object with the exact same set of pixels as object k . We then dilate object k , enlarging it by a few pixels in all directions, and again calculate the probability that the seen or mental image contains an object with the same set of pixels as the dilated object k . We iterate these calculations over a range of sizes for each object, and then plot the probabilities as a function of size for both seen and mental images (Fig. 5). These plots effectively provide object-size tuning functions for vision and imagery. In this preliminary dataset, for simple target images like candle-01 the most likely sizes of seen and imagined objects were the same, although the object-size tuning function for the imagined candle was more broad than the tuning function for the seen candle. For complex target images like cheeseburger-03, the most likely sizes of the imagined objects were larger than the seen objects *except* for the background (which is treated as just another object in this analysis), which was smaller in the mental than the seen images. These preliminary results suggest that imagined foreground objects are large relative to seen foreground objects, while imagined backgrounds are relatively small.

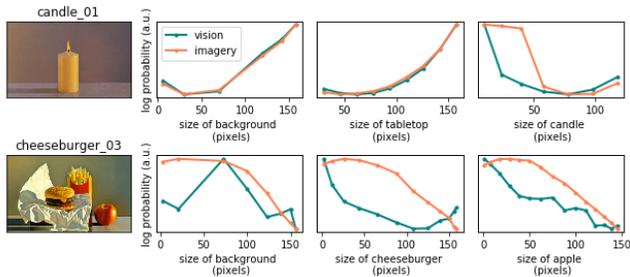


Figure 5: The mental cheeseburger effect: distortions of imagined object size. Top) The target image (left) depicts a candle on a tabletop with a background wall. We used the variational posterior over seen images $q_{vis}(Z)$ and mental $q_{img}(Z)$ images to obtain a distribution over the sizes (in pixels) of the objects in the target image. The background is explicitly treated as an object (second from left). For this target image the distribution over seen (blue) and imagined (orange) background sizes are identical (up to a scaling constant which has here been factored out). The same is true for the table-top (third from left). The most probable sizes of the seen and imagined candle are also identical, even though the size distribution for the imagined candle has more entropy. (Bottom) The target image depicts a cheeseburger, wrapper, fries, and an apple on a tabletop with a background wall. For this more complex target image the distributions over the seen and imagined background (second left), cheeseburger (third from left) and apple (rightmost) are quite different. The seen background occupies more space than the imagined background, while the mental cheeseburger and mental apple are larger than their visual counterparts.

Uncertainty about imagined objects

We obtained point estimates of the hallucination and forgetting rates associated with each subject’s seen or imagined representation of each target image. The hallucination and forgetting rates provide distinct measures of subjects’ uncertainty about seen and mental images. Our preliminary results (Fig. 6) suggest that subjects are far more likely to forget objects that are present in their mental images than to hallucinate them, particularly for complex images. There is much less of either kind of uncertainty about seen objects.

Conclusions

We have presented *viral*, a method that uses subjective reports to infer a distribution over seen and mental images. Given a small amount of noisy data *viral* generates accurate cross-validated predictions of subjective reports. Preliminary results obtained with *viral* support the intuition that as mental images become more complex, the objects in them become distorted and our uncertainty about these objects increases. These effects bound the complexity of mental images, effectively reducing their spatial resolution relative to seen ones. Interestingly, dorsal and parietal visual ar-

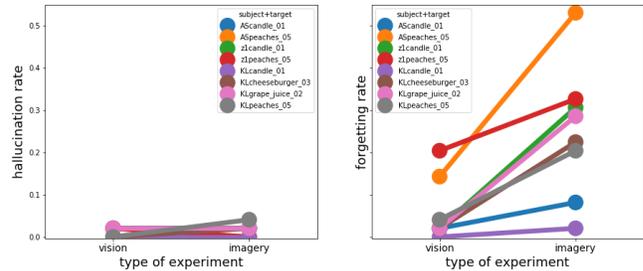


Figure 6: People forget. Hallucination (left) and forgetting (right) rates for all subjects and visual and mental target images. Hallucination rates are negligible; forgetting rates increase with complexity of the target image and are always larger for mental imagery than for vision.

eaas strongly implicated in the generation of mental images (Breedlove et al., 2017) and in the representation of object numerosity (Harvey, Klein, Petridou, & Dumoulin, 2013) maintain a low-resolution representation of the visual world. We speculate that limitations on the complexity of mental images might be determined by limitations on the spatial resolution of representations encoded in these visual brain areas.

Acknowledgments

Funding for this work was provided by grant NEI R01 EY023384 to TN. We acknowledge Katherine Lynam for data collection, and Ghislain St-Yves for helpful discussion. All paintings by David Ligare. candle-01: “Still life with candle”, 1999; grape-juice-02: “Still life with grape juice and sandwiches (Xenia), 1994; cheeseburger-03: “Still life with burger, fries and apple”, 1995; peaches-05: “Still life with peaches and water (Aparchai)”, 2005.

References

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2016). Variational Inference: A Review for Statisticians. *arXiv:1601.00670 [cs, stat]*.

Breedlove, J., Naselaris, T., & St-Yves, G. (2017). A theory of mental imagery. In *Conference on cognitive computational neuroscience*.

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic Representation of Numerosity in the Human Parietal Cortex. *Science*, 341(6150), 1123–1126.

MacKisack, M., Aldworth, S., Macpherson, F., Onians, J., Winlove, C., & Zeman, A. (2016). On Picturing a Candle: The Prehistory of Imagery Science. *Frontiers in Psychology*, 7.

Podgorny, P., & Shepard, R. N. (1978). Functional representations common to visual perception and imagination. *Journal of experimental psychology. Human perception and performance*, 4(1), 21–35.