Pinpointing the neural signatures of single-exposure visual recognition memory

Vahid Mehrpour[a], Travis Meyer[a], Eero P. Simoncelli[b], and Nicole C. Rust[a,1]

[a]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104; and [b]Flatiron Institute, Simons Foundation and Center for Neural Science, New York University, New York, NY 10010

Edited by Winrich A. Freiwald, Rockefeller University, New York, NY, and accepted by Editorial Board Member Thomas D. Albright March 25, 2021 (received for review October 16, 2020)

Memories of the images that we have seen are thought to be reflected in the reduction of neural responses in high-level visual areas such as inferotemporal (IT) cortex, a phenomenon known as repetition suppression (RS). We challenged this hypothesis with a task that required rhesus monkeys to report whether images were novel or repeated while ignoring variations in contrast, a stimulus attribute that is also known to modulate the overall IT response. The monkeys' behavior was largely contrast invariant, contrary to the predictions of an RS-inspired decoder, which could not distinguish responses to images that are repeated from those that are of lower contrast. However, the monkeys' behavioral patterns were well predicted by a linearly decodable variant in which the total spike count was corrected for contrast modulation. These results suggest that the IT neural activity pattern that best aligns with single-exposure visual recognition memory behavior is not RS but rather sensory referenced suppression: reductions in IT population response magnitude, corrected for sensory modulation.

recognition memory | repetition suppression | contrast | population decoding | familiarity

Under the right conditions, we are very good at remembering the images that we have seen: we can remember thousands of images after viewing each only once and only for a few seconds (1, 2). How our brains support this remarkable ability, often called "visual recognition memory" (3), is not well understood. The most prominent proposal to date suggests that memories about whether images have been encountered before are signaled in high-level visual brain areas such as inferotemporal cortex (IT) and perirhinal cortex via adaptation-like reductions of the population response to repeated as compared to novel stimuli, a phenomenon referred to as repetition suppression (RS) (4–9). Repetition suppression exhibits the primary attributes needed to account for the vast capacity of single-exposure visual memory behavior: response decrements in subsequent exposures are selective for image identity (even after viewing an extensive sequence of other images), and last for several minutes to hours (5, 6, 10). RS has also been shown to account for behavior in an image recognition memory task: a linear decoder with positive weights can predict single-exposure visual recognition memory behavior from neural responses in IT cortex (10).

Despite the fact that the RS hypothesis is consistent with available evidence, it seems likely to be too simplistic an explanation for visual recognition memory encoding. In particular, it is well known that sensory neurons such as those of IT cortex are modulated not only by image memory, but also by stimulus properties such as image contrast (11). It is thus unclear whether and how these stimulus-induced effects interfere with judgments of whether images are novel or have been encountered before, and if they do not, how image memory can be decoded from neural responses in a way that disambiguates it from changes in these stimulus properties. To investigate this, we measured behavioral and neural responses of monkeys trained to report whether images were novel or repeated while disregarding image contrast (Fig. 1*A*).

Results

The Contrast-Invariant Visual Memory Task. Monkeys viewed sequences of grayscale images, each presented for 500 ms, and each presented exactly twice (initially novel, then repeated). Novel and repeated images were presented with equal probability in all possible combinations of high (H) and low (L) contrasts, including (novel, repeated): HH, LL, HL, and LH. We refer to the former two cases as the "same-contrast" conditions and the latter two as the "mixed-contrast" conditions (Fig. 1*B*). Monkeys were trained to report, on each trial, whether the observed image was novel or repeated, while disregarding image contrast (Fig. 1*A*). Here we refer to the change in state between novel and repeated trials as "image memory."

After training, the monkeys were largely able to disambiguate image memory from changes in image contrast: they performed equally well for both same-contrast conditions, and they were only modestly impaired for the mixed-contrast conditions (Fig. 1*C*). A two-way ANOVA applied to the behavioral data revealed a statistically significant interaction between memory and contrast (combined data: $P = 1.74 \times 10^{-5}$; monkey 1: $P = 0.044$; monkey 2: $P = 3.62 \times 10^{-4}$), in the absence of a significant modulation by contrast (combined data $P = 0.750$; monkey 1: $P = 0.355$; monkey 2: $P = 0.090$) and in the presence of a significant modulation by memory (combined data $P = 7.82 \times 10^{-7}$; monkey 1: $P = 0.001$; monkey 2: $P = 1.48 \times 10^{-5}$). Because an interaction between memory and contrast can take on many different behavioral patterns, we quantified the degree to which this interaction reflected systematic confusion between memory and contrast by comparing the monkeys' behavioral patterns with the pattern that would be expected from systematic contrast confusions (shown in the *Inset* of Fig. 1*C*, e.g., disproportionately reporting that low contrast images

Significance

Humans have a remarkable ability to remember images they have seen, even after seeing thousands, each only once and for a few seconds. One important step toward understanding how the primate brain supports this remarkable form of memory involves pinpointing the neural activity patterns that enable image memory behavior. This paper presents evidence this neural activity pattern is sensory referenced suppression: reductions in population response magnitude, corrected for sensory modulation.

Author contributions: V.M., T.M., E.P.S., and N.C.R. designed research; V.M., T.M., and N.C.R. performed research; V.M., E.P.S., and N.C.R. contributed new reagents/analytic tools; V.M. and N.C.R. analyzed data; and V.M., E.P.S., and N.C.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. W.A.F. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: nrust@sas.upenn.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2021660118/-/DCSupplemental.

Published April 26, 2021.

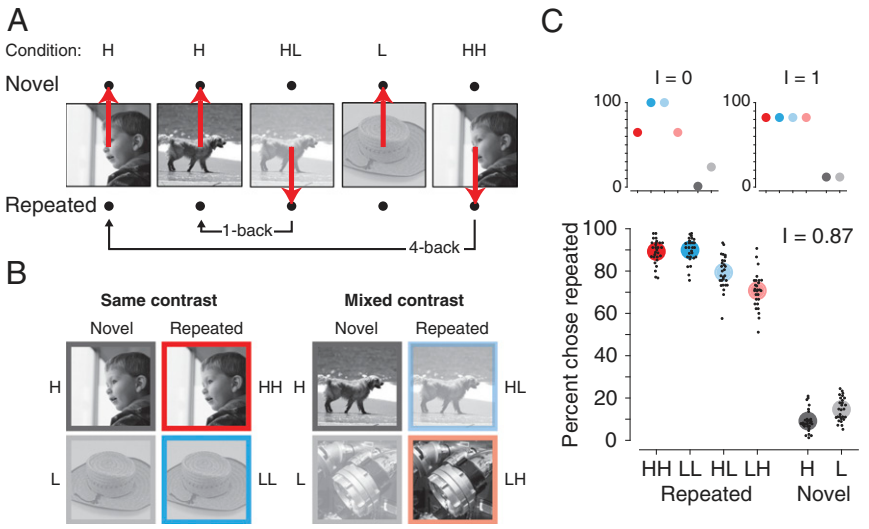

Fig. 1. Visual memory behavior. (*A*) The contrast-invariant, single-exposure visual memory task. The monkeys viewed a sequence of images and reported whether they were novel (never seen before) or repeated (seen exactly once) while ignoring randomized changes in contrast. Monkeys were trained to saccade to one of two response targets to indicate their choice (red arrows). Images were repeated with a randomly chosen delay between the first and repeated presentation ("n-back"). (*B*) Images were displayed at one of two contrast levels, yielding two conditions for novel images, high (H) and low (L), and four conditions for repeated images: HH (repeated H preceded by novel H), LL (repeated L preceded by novel L), HL (repeated L preceded by novel H), and LH (repeated H preceded by novel L). The four repeated conditions were organized into same-contrast and mixed-contrast groups depending on whether the initial and repeated presentations were at the same or different contrasts, respectively. (*C*) Behavioral performance for the data pooled across monkeys in the task, where small black dots indicate average performance for an individual session and large colored dots indicate the average performance across sessions. A measure of contrast invariance, I, was computed as the ratio of the variance across contrast conditions and the variance with respect to the maximally contrast-modulated pattern after taking overall performance into account, subtracted from 1 (*SI Appendix, SI Methods*). *Insets* illustrate the expected behavioral pattern with minimal (I = 0) and maximal (I = 1) contrast invariance.

are repeated). This index (I), was constrained to range 0 to 1, where 1 indicates a behavioral pattern that is perfectly contrast invariant and 0 indicates maximal contrast confusion after taking into account the monkeys' overall performance (Fig. 1*C*, *Insets*). Behavioral contrast invariance values were high (combined data: 0.87; monkey1: 0.95, monkey2: 0.84; *SI Appendix*, Fig. S1), indicating that the monkeys were able to judge image memory while largely (albeit imperfectly) avoiding systematic confusion with changes in image contrast. As detailed in the next section, these behavioral data challenge existing proposals that IT repetition suppression is the neural signal underlying image memory, as these proposals predict that memory and contrast will be systematically confused.

RS and Optimally Weighted Linear Decoders Fail to Predict Behavior. As the monkeys performed the task, we recorded neural responses in IT. Because accurate estimates of population response magnitude require many hundreds of units, data were concatenated across sessions into a larger pseudopopulation in a manner that combined trials within the same experimental condition (*SI Appendix*, *SI Methods*). Spikes were counted in a window starting 100 ms after stimulus onset (to allow for the latency of visual signals arriving in IT) and ending 400 ms later, at the termination of the image viewing period. The resulting pseudopopulation contained the responses of 856 units to 180 images each presented twice, and distributed evenly (and randomly) within the four conditions (i.e., 45 images for each of HH, LH, HL, and LL). As an initial, summary analysis of the IT neural data, we quantified the magnitudes of both memory and contrast modulations as proportional reductions in the overall grand mean firing rate from the novel H condition to the repeated HH and novel L conditions for memory and contrast, respectively. Modulations for memory and contrast were 6% versus 3% when applied to the raw responses, and 13% versus 7% after subtracting out the prestimulus onset baseline.

Next, we assessed the hypothesis that RS of IT responses can explain visual memory behavior. We instantiated this hypothesis with a total spike count decoder, in which image memory was determined by comparing the total spike count with a threshold. We quantified the degree of alignment between neural predictions of behavioral patterns and the monkeys' actual behavior with a measure termed "prediction quality (PQ)." PQ was computed from the mean squared error between the actual behavioral patterns and best-fitting neural prediction of behavior (*SI Appendix*, *SI Methods*). The upper bound for our measure, PQ = 1, reflects a neural prediction that perfectly replicates the actual behavioral pattern, and PQ = 0 reflects the worst possible predicted behavioral pattern that was matched in overall performance (e.g., a pattern that was modulated entirely by changes in contrast, analogous to the *Insets* in Fig. 1*C*). This RS decoder produced a behavioral prediction that reflected confusions between changes in image memory with changes in contrast, for both repeated as well as novel images (compare with the *Insets* in Fig. 1*C*) and low PQ (PQ_{RS} = 0.40; Fig. 2*A*). A control analysis confirmed that the predicted behavioral patterns on repeated trials were not a consequence of misclassifications of those images when they were presented as novel (*SI Appendix*, Fig. S2), consistent with the interpretation that these behavioral patterns reflect confusion with contrast as opposed to other factors.

The RS decoder is a linear decoder with uniform weighting over the neural population, so we wondered whether more carefully chosen weights might yield a linear decoder that could match the behavioral responses. Specifically, an optimally weighted linear decoder was previously shown to be effective at aligning IT neural responses with visual memory behavior in the absence of contrast modulation (10). We used this same Fisher linear discriminant, computed assuming independence of neural responses, that weights each unit proportional to its memory discriminability, d′ (iFLD) (*SI Appendix*, *SI Methods*). The iFLD differs from RS in that it weights each unit according to the amount of task-relevant

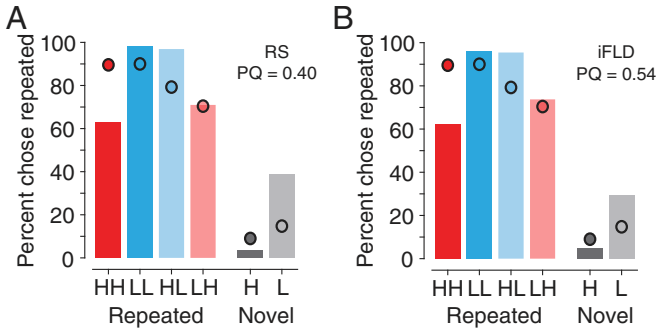

Fig. 2. Traditional linear decoders confuse image memory and contrast and fail to map IT neural responses to behavior. Each panel reflects the monkeys' actual behavioral patterns (dots) along with the predictions of a linear decoder applied to the recorded neural population (bars). (*A*) Total spike count decoder, motivated by RS. (*B*) Optimally weighted linear decoder, iFLD. Prediction quality (PQ) quantifies similarity between the neural predictions of behavior and the monkeys' actual behavioral patterns (see text).

information that it carries, and these weightings are signed: any units that exhibit repetition enhancement (on average) would be appropriately combined (with opposite sign) with units that exhibit repetition suppression. Despite the fact that this decoder is optimized to extract image memory information while disregarding contrast, we found that the iFLD also confused changes in image memory with changes in image contrast, and behavioral predictions were only slightly improved relative to RS (PQ_{iFLD} = 0.54; Fig. 2*B*). Poor behavioral predictions for RS and iFLD were replicated for each monkey individually (*SI Appendix*, Fig. S3; monkey 1: $PQ_{RS} = 0.61$, $PQ_{iFLD} = 0.66$; monkey 2: $PQ_{RS} = 0.19$, and $PQ_{iFLD} = 0.53$). We return to examine the underlying reasons for this failure below; see Fig. 5.

Sensory Referenced Suppression Is a Good Predictor of Behavior. We wondered whether the monkeys' behavioral patterns could be explained by an alternative linear decoder applied to the IT population responses. Given the substantial evidence in support of the repetition suppression hypothesis, we reasoned that the brain might be acting on a variant of this neural signature in which it corrects for the ambiguities in total spike count that are introduced by changes in contrast. Because this hypothetical decoding scheme operates by estimating and correcting for modulations in the total spike count due to variations in memory-irrelevant sensory attributes, we refer to this hypothesis as "sensory referenced suppression (SRS)."

What would be required for SRS to be an effective account of the mapping of IT neural signals to behavior, if such a decoding scheme were restricted to act only on the IT population response? The fact that the total spike count is affected by contrast implies that the optimal linear decoder for contrast must at least partially overlap with (i.e., be nonorthogonal to) the RS decoding axis (a vector of ones, representing equal weights for each unit). We found that this was indeed the case: when applied to the pooled data, an optimized decoding vector for contrast lies in a direction 69° from the total spike count vector (labeled RS), indicating that information about contrast was largely nonoverlapping but not orthogonal to RS (Fig. 3*A*). Next, we considered the family of linear discrimination vectors that live on the two-dimensional (2D) plane defined by RS and the contrast decoder. On this plane, we defined the direction of the RS decoder (with no contrast correction) as 0° (see Fig. 3*A*, *Top Inset*: RS). A decoding vector that is rotated clockwise from the RS decoder (away from the contrast decoder) in this plane is less affected by stimulus contrast. Rotation toward the contrast decoder exacerbates contrast modulation in the predicted behavioral patterns. Within this family of linear decoding schemes, we define SRS as the decoder that is orthogonal

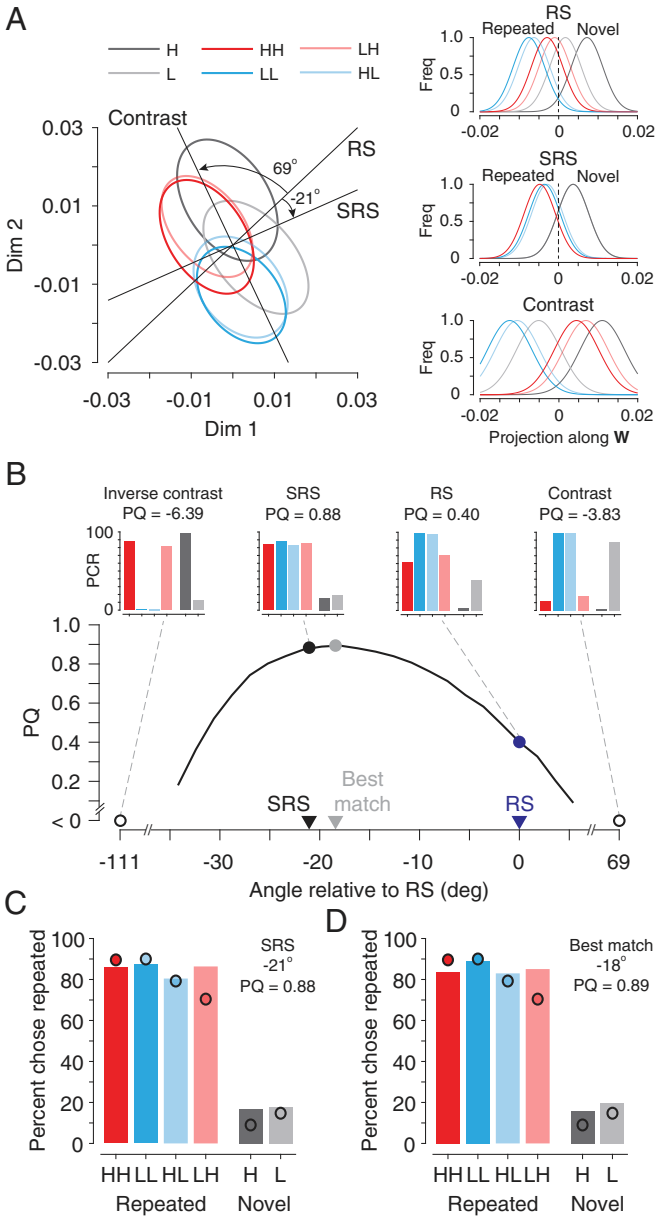

Fig. 3. Neural predictions of behavior for a family of weighted linear decoders that include RS, an optimized contrast decoder, and SRS. (*A*) Projections of IT neural response distributions for all six stimulus conditions onto the 2D plane defined by weight vectors for the total spike count vector (RS, which uses a weight vector of all ones) and for a linear decoder optimized for contrast ("Contrast"). Ellipses depict 95% probability intervals for 2D histograms of the projection of neural responses onto this plane (*SI Appendix, SI Methods*). *Insets* show 1D histograms of the projections of the distributions onto the three linear decoders. (*B*) The quality of the neural predictions of monkeys' behavioral patterns (PQ) for the family of linear decoders that lie within the plane. Negative PQ values reflect predicted behavioral patterns that could not be rescaled to match overall performance because one or more entries were pinned at saturation (e.g., as a consequence of extreme contrast modulation). Each decoder corresponds to a rotation of the total spike count decoder, or equivalently, the weighted combination of the total spike count decoder and the contrast decoder. Markers indicate: SRS (black), which has minimal contrast sensitivity (i.e., orthogonal to the contrast axis); RS (blue), the total spike count decoder with no contrast correction; and the best behavioral match (maximal PQ in gray). *Insets* above depict the corresponding neural predictions of behavior. (*C* and *D*) The alignment of the monkeys' actual behavioral patterns (dots) and the neural predictions of behavior (bars) for (*C*) SRS and (*D*) the decoder with the best behavioral match.

to (i.e., 90° from) the contrast decoder, and consequently minimizes contrast modulation in the neural prediction of behavioral patterns. The SRS was −21° from RS for the data pooled across both monkeys (Fig. 3*B*) and −23° and −18° for individual animals (*SI Appendix*, Fig. S3).

The SRS linear decoder provides good predictions of the monkeys' behavioral patterns, both for the pooled data (PQ_{SRS} = 0.88; Fig. 3*C*), and for each monkey individually (monkey 1: PQ_{SRS} = 0.87; monkey 2: PQ_{SRS} = 0.93; *SI Appendix*, Fig. S4). It also provided a much better prediction of behavior than RS or the iFLD (pooled data: PQ_{RS} = 0.40 and PQ_{iFLD} = 0.54; monkey 1: PQ_{RS} = 0.61 and PQ_{iFLD} = 0.66; monkey 2: PQ_{RS} = 0.19 and PQ_{iFLD} = 0.53). These results suggest that SRS provides a considerably better description of the relationship between IT neural activity and behavior than RS or iFLD under the challenge of sensory-induced variations in population response magnitude (i.e., contrast modulation).

The SRS decoder can be thought of as isolating image memory information by correcting the IT population response for contrast modulation. We also considered a variant decoding scheme in which the contrast correction was applied by estimating and then subtracting contrast modulation from the RS decoder response (*SI Appendix*, *SI Methods*). The behavioral pattern predicted by this decoder reflected a higher degree of contrast modulation and was less well aligned with behavior (PQ = 0.75; *SI Appendix*, Fig. S6) than SRS (PQ_{SRS} = 0.88; Fig. 3*C*).

The results presented thus far were obtained using response values averaged over a temporal window from 100 to 500 ms following stimulus onset. We found similar results when the analysis was applied to shorter time windows placed at different positions relative to stimulus onset. These analyses revealed that contrast modulation preceded memory modulations in IT and was reflected throughout the entire viewing period following the initial latency delay (*SI Appendix*, Fig. S5 *A* and *B*). Consequently, PQ_{SRS} rose and then remained high until the end of the viewing period and PQ_{SRS} was higher than PQ_{RS} for all time windows (*SI Appendix*, Fig. S5*B*).

The SRS Decoder Had Better Image Memory Performance than RS. To better understand how memory and contrast were reflected in IT during these experiments, we shifted our focus away from the alignment between decoding predictions and behavior and toward overall performance for decoding image memory. These issues are best conceptualized by considering discriminability, rather than percent correct, as a measure of performance. Discriminability (often labeled d′ in the perceptual literature) is defined as the ratio of the difference between the means of the novel and repeated distributions divided by the square root of the average variance of those distributions (Fig. 4*A*). In our experiments, the variance of each distribution can be further decomposed into two components: 1) modulations within each distribution by contrast (Fig. 4*A*, "Contrast Var") and 2) combined modulations arising from image identity and trial-to-trial variability (which cannot be dissociated, due to the single-trial nature of these experiments; Fig. 4*A*, "ID Var") (*SI Appendix*, *SI Methods*).

We found that, in addition to being a better predictor of behavior (Fig. 3*B*), the SRS decoder also had higher image memory performance than RS (Fig. 4*B*). This occurred despite the fact that novel and repeated means were actually closer together in the SRS direction than the RS direction (Fig. 4*C*). These decreases in mean separation were offset by decreases in variance (denominator of d′), plotted in Fig. 4*D*. These decreases in variance could in turn be attributed entirely to the elimination of contrast modulation. In sum, the superior performance of SRS resulted from novel and repeated distributions whose means were slightly closer together, but whose variances decreased even more as a consequence of eliminating contrast modulation along the SRS linear decoding axis.

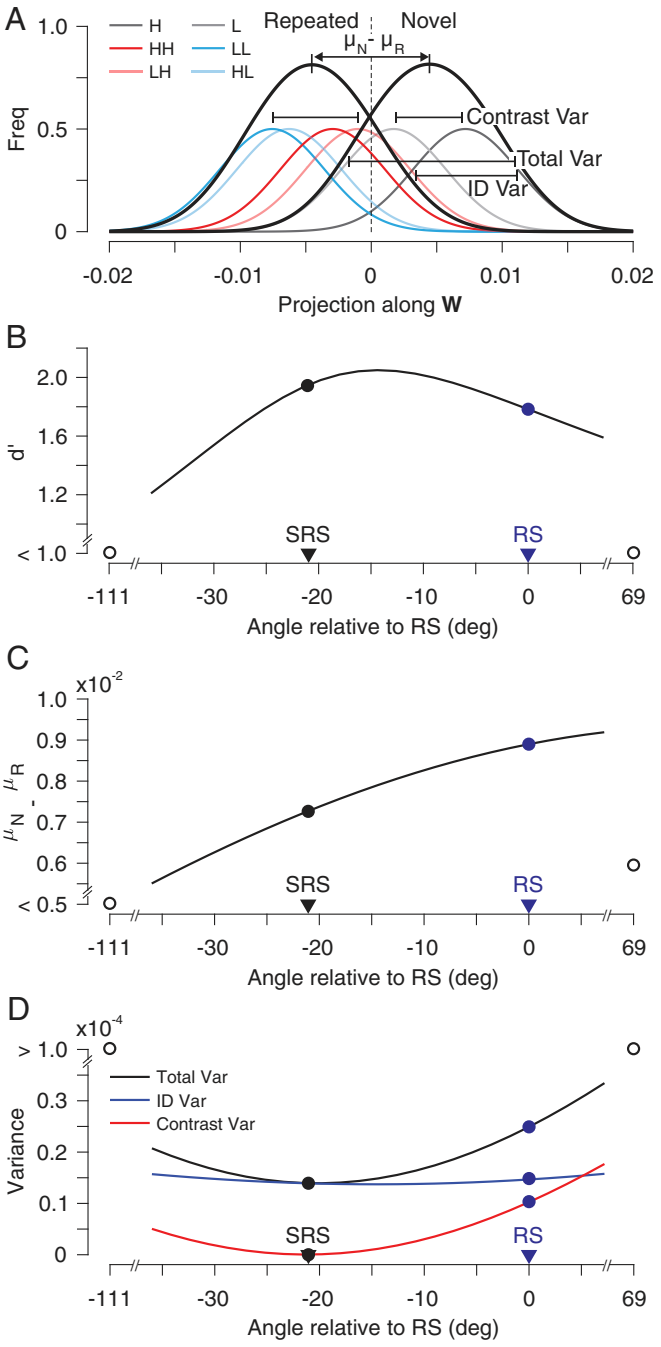

Fig. 4. The population geometry impacting overall image memory performance for SRS and RS. (*A*) Schematic of linear decoder performance, computed as d′, for this task. Shown are 1D histograms of the projection of the IT population responses onto a linear decoding axis **W**. Discriminability for image memory (d′) is computed as the difference between the means of novel and repeated distributions ($\mu_N - \mu_R$) divided by the square root of the average total variance (Total Var). (*B*) d′ as a function of angle on the 2D plane defined in Fig. 3*A*. (*C* and *D*) Decomposition of d′: (*C*) the numerator (difference between means), and (*D*) the square of the denominator, the total variance (Total Var), further broken down into the variance due to image identity and trial-to-trial variability (ID Var) and contrast modulation (Contrast Var) (*SI Appendix*, *SI Methods*). In *B*–*D*, open circles at the *Right* side of each graph indicate the values for projections along the contrast decoder.

Relationship between SRS and iFLD Decoders. The results presented above demonstrate that while the largely contrast-invariant patterns reflected in the monkeys' behavior are consistent with the

SRS decoder (Fig. 3*C*), a linear decoder optimized for image memory on our data (the iFLD) confuses image memory with changes in image contrast (Fig. 2*B*). What do these differences imply about the geometry by which image memory and contrast are reflected in IT? To address these questions, we turned to simulations, where issues about population geometry can be investigated absent the constraints imposed by finite samples. To perform these simulations, we began by fitting a model to each single unit that we recorded. For each IT unit, the distribution of the visually evoked firing rate response over stimuli was modeled by an exponential function (12); image memory and contrast were modeled as multiplicative modulations of the visually evoked response; and the trial-to-trial distribution of spike counts was modeled as an independent Poisson process (*SI Appendix*, *SI Methods*). The four parameters fit for each unit included: 1) mean firing rate (the mean of the exponential), 2) the visually evoked tuning bandwidth, 3) image memory sensitivity, and 4) contrast sensitivity (*SI Appendix*, *SI Methods*). We found that "synthetic" data from the resulting model population recapitulated all aspects of the physiological data that we have highlighted thus far, including contrast modulation in the RS predictions (*SI Appendix*, Fig. S7*A*, *Top Inset*), contrast-invariant SRS (*SI Appendix*, Fig. S7*A*, *Middle Inset*), and overall d′ that was higher for SRS than RS as a consequence of eliminating contrast modulation (*SI Appendix*, Fig. S7 *B–D*).

Next, to understand the relationship between SRS and the iFLD, and why the iFLD did not exhibit contrast invariance, we performed a set of analysis similar to those described for Figs. 3 and 4 but within the plane spanned by SRS and iFLD (Fig. 5*A*). The iFLD is optimal (under the assumption of Gaussian-distributed independent response), and indeed has higher discrimination performance than SRS (Fig. 5*B*). Increased d′ for iFLD over SRS resulted from both an increase in the distance between the means of distributions for novel and repeated images (i.e., the d′ numerator; Fig. 5*C*) as well as a decrease in the variance of distributions for novel and repeated images (i.e., the d′ denominator; Fig. 5*D*). Intriguingly, the overall reduction in total variance along the iFLD axis relative to SRS resulted from an increase in contrast modulation that were offset by a larger decrease in identity modulation relative to SRS (Fig. 5*D*). This was because identity modulation and contrast modulation were anticorrelated on this plane: decreases in one (e.g., identity modulation) were accompanied by increases in the other (e.g., contrast modulation; Fig. 5*D*). In other words, the iFLD failed to predict contrast invariance in behavioral patterns because it could achieve higher image memory performance by reducing identity variance, which was anticorrelated with contrast.

To complement the 2D plots presented in Fig. 4 (RS and SRS) and Fig. 5 (SRS and iFLD), *SI Appendix*, Fig. S8 depicts memory decoding performance in the three-dimensional (3D) space defined by SRS, RS, and iFLD, for both the real and synthetic data.

Discussion

Humans and nonhuman primates have a remarkable ability to remember the images that they have seen (1, 2, 5, 6, 13). It has been suggested that image memory is signaled in high-level visual brain areas such as IT via changes in population response magnitude, known as repetition suppression (RS) (4–9). We have challenged this explanation, by examining neural and behavioral memory responses while independently manipulating image contrast. IT population response was modulated by contrast, but monkeys' behavioral reports of image memory were largely invariant to changes in image contrast (Fig. 1), inconsistent with the RS hypothesis (Fig. 2*A*). Behavioral invariance also could not be reconciled with our previous work, which proposed that image memory is linearly decoded from IT by weighting each neuron proportional to the amount and sign of memory-relevant information that it carries (10) (Fig. 2*B*). However, the monkeys' behavioral patterns were linearly decodable from IT (Fig. 3*C*), using a linear decoder that corrects the total spike count decoder by

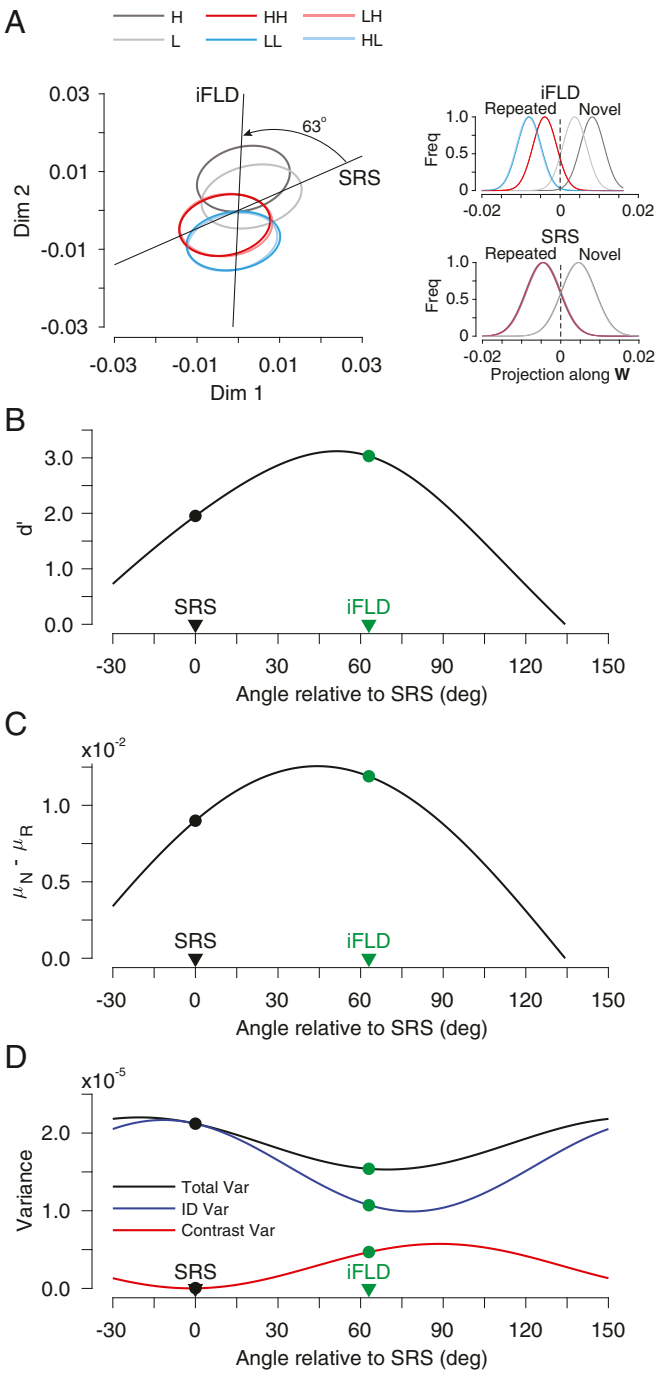

Fig. 5. The population geometry impacting overall performance for SRS and iFLD. To explore population geometry absent the constraints imposed by limited samples, a model was fit to each unit and model parameters were used to create synthetic data. (*A*) Projections of the synthetic data onto the 2D plane defined by SRS and a linear decoder optimized for memory, "iFLD." Ellipses depict 95% probability intervals for 2D histograms of the projection of neural responses onto this plane. *Insets* show 1D histograms of the projections onto each linear axis. (*B*) d′ as a function of angle relative to SRS on the 2D plane defined in *A*. (*C* and *D*) Decomposition of d′ into (*C*) its numerator, the difference between the means of the novel and repeated distributions and (*D*) the square of its denominator, the total variance (Total Var), further broken down into the variance due to image identity and trial-to-trial variability (ID Var) and contrast modulation (Contrast Var). In *B–D* values corresponding to SRS and iFLD are labeled by black and green markers, respectively.

eliminating its contrast dependence. We refer to this linear decoding scheme as sensory referenced suppression (SRS), because it can be understood as estimating image memory from the total spike count after correcting for sensory modulation (Fig. 3*A*).

The hypothesis that image memory is encoded in high-level visual cortex as RS has a mixed history, with some studies finding support for this hypothesis (6, 8, 10, 14, 15) and others finding evidence against it (16, 17). In an earlier study, we reported that some instantiations of RS decoders were good predictors of the rates of behavioral remembering and forgetting when tested with randomly selected images, under the assumption that all novel images evoke the same magnitude population response from IT (10). The work we present here suggests that modifications of RS are required to account for single-exposure visual memory behavior when factors other than image memory modulate the magnitude of the population response. That is, we find that when novel images evoke lower firing rates due to changes in contrast (e.g., L vs. H), an RS classifier confuses those differences in firing rate due to contrast for differences in image memory, and predicts higher memory performance when those images are repeated (e.g., LL versus HH). These predicted patterns are at odds with the actual behavioral patterns of the monkeys, who exhibit similar performance for the LL and HH conditions. SRS resolves this discrepancy by removing the contrast dependency from the RS memory encoding scheme.

We also considered a variant linear decoder in which the contrast correction was estimated based on the responses to novel images and then subtracted from the RS decoder response (*SI Appendix*, Fig. S6). This decoder did not predict behavior as well as SRS. Under the assumption that memory and contrast both act by multiplicatively modulating the responses of individual IT neurons (which provides a good description of our data, as shown in *SI Appendix*, Fig. S7), the operations performed by this variant decoder are not expected to perfectly eliminate contrast modulation along the memory axis. As a simple numerical demonstration, assume that contrast and memory modulations are both multiplicative, reducing responses by 10% and 20%, respectively, and the responses for four illustrative conditions are thus H = 1, L = 0.9, HH = 0.8, LL = 0.72. While the goal of the contrast correction is to perfectly align the responses for novel (H and L) and repeated (HH and LL) conditions, a subtractive contrast correction (H − L = 0.1) applied to the low contrast conditions L and LL would successfully align the novel conditions but imperfectly align the repeated conditions (H = 1, L = 1, HH = 0.8, LL = 0.82). Rather, when contrast acts multiplicatively, division (rather than subtraction) is the appropriate way to correct for it. However, a divisive correction would require a nonlinear decoding scheme. What is surprising about SRS is that it accomplishes the same goal as this hypothetical nonlinear decoder with a purely linear decoding scheme (by orthogonalizing the total spike count axis against the contrast axis, thereby eliminating it).

A number of factors other than contrast are known to modulate the IT population response, including stimulus attributes such as object size (11), a diverse set of stimulus attributes that contribute to image memorability (18–20), and external factors such as surprise (21, 22) and attention (8, 23). The SRS decoding scheme that we have proposed could, in principle, provide a mechanism for the brain to disambiguate image memory-induced changes in IT population response magnitude from changes due to the combination of all of these other factors. In principle, all that is required to generalize the SRS decoding scheme is for the effect of these other factors to be at least partially nonoverlapping with that of image memory (e.g., if individual units have heterogeneous sensitivities for these). In this case, a linear decoder can be constructed by eliminating each of these factors in turn from the RS decoder, analogous to the successive orthogonalizations performed in the Gram–Schmidt procedure. Future work will be required to determine how such a decoding scheme might be learned by the brain. Learning algorithms for decoding often presume the existence of labels (e.g., "novel" versus "repeated"), with the general idea that such feedback ("supervision") is available while a subject is actively learning a task. It is less clear how SRS could be learned without such supervision (in the task of our experiments, animals only receive reinforcement feedback regarding correct/incorrect choices, not specific feedback regarding contrast).

What is the origin of the IT magnitude variation that aligns with single-exposure visual recognition memory behavior? RS is found at all stages of visual processing from the retina to IT, and it strengthens in both magnitude and duration as one ascends the visual cortical hierarchy (24). Consequently, a hierarchical cascade of feed-forward, adaptation-like mechanisms may underlie RS measured in IT (25). There are also indications that RS within IT may arise from changes in synaptic weights between recurrently connected units within IT itself (25, 26). Finally, a component of RS in IT is likely to be fed back to IT from higher brain areas such as perirhinal cortex or hippocampus. While the assertion that top-down processing contributes to RS in high-level visual cortex has been controversial (25, 27–29), recent evidence from a patient with medial temporal lobe (MTL) damage supports a role for feedback from MTL structures to RS in high-level visual cortex (30). Within one MTL structure, the hippocampus, single-exposure recognition memory behavior has been linked with RS (31, 32) as well as synchronizations between gamma oscillations and spikes (33). However, because these evaluations were not made in a manner that challenges RS with other factors that affect response magnitude, additional work will be required to determine whether SRS is a better description than RS of the neural signatures that reflect single-exposure visual recognition memory behavior in MTL structures that lie downstream from IT. What is clear is that pinpointing the neural signatures that align with single-exposure visual memory behavior is a good first step toward understanding how the primate brain manages to remember images so remarkably well.

Methods

Experiments were performed on two adult male rhesus macaque monkeys. All procedures were performed in accordance with the guidelines of the University of Pennsylvania Institutional Animal Care and Use Committee. In these experiments, electrophysiological recordings were made from neurons in IT cortex as the monkeys performed a single-exposure visual memory task. The task, electrophysiological recordings, and data analyses are described in *SI Appendix, SI Methods*.

Data Availability. Data have been deposited in Dryad (https://doi.org/10.5061/dryad.gmsbcc2mh).

ACKNOWLEDGMENTS. Recent work in neuroscience and related fields has identified citation biases whereby work from women and minorities are undercited relative to other papers in the field (34–36). In crafting this manuscript, we sought to proactively consider citation bias. Following ref. 34, the gender balance of citations was quantified based on the first names of the first/last authors using open source code (37). Excluding self-citations, this article contains 58.3% man/man, 16.7% man/woman, 19.4% woman/man, and 5.6% woman/woman citations. For comparison, proportions estimated from articles in the five top neuroscience journals (as reported in ref. 34) are 58.4% man/man, 9.4% man/woman, 25.5% woman/man, and 6.7% woman/woman. This work was supported by the Simons Foundation (Simons Collaboration on the Global Brain award 543033 to N.C.R. and 543047 to E.P.S.), the National Eye Institute of the NIH (award R01EY020851 to N.C.R.), the NSF (CAREER award 1265480 to N.C.R.), and the Howard Hughes Medical Institute (investigatorship to E.P.S.).

1. L. Standing, Learning 10,000 pictures. *Q. J. Exp. Psychol.* **25**, 207–222 (1973).
2. T. F. Brady, T. Konkle, G. A. Alvarez, A. Oliva, Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14325–14329 (2008).
3. J. B. Rich, "Recognition memory" in *Encyclopedia of Clinical Neuropsychology*, J. Kreutzer, J. DeLuca, B. Caplan, Eds. (Springer, New York, NY, 2011), p. 156.
4. F. L. Fahy, I. P. Riches, M. W. Brown, Neuronal activity related to visual recognition memory: Long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. *Exp. Brain Res.* **96**, 457–472 (1993).
5. L. Li, E. K. Miller, R. Desimone, The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* **69**, 1918–1929 (1993).

6. J. Z. Xiang, M. W. Brown, Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology* **37**, 657–676 (1998).
7. R. Desimone, Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13494–13499 (1996).
8. E. K. Miller, L. Li, R. Desimone, A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* **254**, 1377–1379 (1991).
9. I. P. Riches, F. A. Wilson, M. W. Brown, The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. *J. Neurosci.* **11**, 1763–1779 (1991).
10. T. Meyer, N. C. Rust, Single-exposure visual memory judgments are reflected in inferotemporal cortex. *eLife* **7**, e32259 (2018).
11. D. Zoccolan, M. Kouh, T. Poggio, J. J. DiCarlo, Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J. Neurosci.* **27**, 12292–12307 (2007).
12. N. C. Rust, J. J. DiCarlo, Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream. *J. Neurosci.* **32**, 10170–10182 (2012).
13. J. L. Ringo, R. W. Doty, A macaque remembers pictures briefly viewed six months earlier. *Behav. Brain Res.* **18**, 289–294 (1985).
14. B. D. Gonsalves, I. Kahn, T. Curran, K. A. Norman, A. D. Wagner, Memory strength and repetition suppression: Multimodal imaging of medial temporal cortical contributions to recognition. *Neuron* **47**, 751–761 (2005).
15. N. B. Turk-Browne, D. J. Yi, M. M. Chun, Linking implicit and explicit memory: Common encoding factors and shared representations. *Neuron* **49**, 917–927 (2006).
16. E. J. Ward, M. M. Chun, B. A. Kuhl, Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. *J. Neurosci.* **33**, 14749–14757 (2013).
17. G. Xue *et al*., Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *J. Cogn. Neurosci.* **23**, 1624–1633 (2011).
18. A. Jaegle *et al*., Population response magnitude variation in inferotemporal cortex predicts image memorability. *eLife* **8**, e47596 (2019).
19. P. Isola, Jianxiong Xiao, D. Parikh, A. Torralba, A. Oliva, What makes a photograph memorable? *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 1469–1482 (2014).
20. W. A. Bainbridge, P. Isola, A. Oliva, The intrinsic memorability of face photographs. *J. Exp. Psychol. Gen.* **142**, 1323–1334 (2013).
21. T. Meyer, C. R. Olson, Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19401–19406 (2011).
22. C. M. Schwiedrzik, W. A. Freiwald, High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron* **96**, 89–97.e4 (2017).
23. N. Roth, N. C. Rust, Inferotemporal cortex multiplexes behaviorally-relevant target match signals and visual representations in a manner that minimizes their interference. *PLoS One* **13**, e0200528 (2018).
24. J. Zhou, N. C. Benson, K. N. Kay, J. Winawer, Compressive temporal summation in human visual cortex. *J. Neurosci.* **38**, 691–709 (2018).
25. R. Vogels, Sources of adaptation of inferior temporal cortical responses. *Cortex* **80**, 185–195 (2016).
26. S. Lim *et al*., Inferring learning rules from distributions of firing rates in cortical neurons. *Nat. Neurosci.* **18**, 1804–1810 (2015).
27. C. Summerfield, E. H. Trittschuh, J. M. Monti, M. M. Mesulam, T. Egner, Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* **11**, 1004–1006 (2008).
28. M. Grotheer, G. Kovács, Repetition probability effects depend on prior experiences. *J. Neurosci.* **34**, 6640–6646 (2014).
29. K. Vinken, H. P. Op de Beeck, R. Vogels, Face repetition probability does not affect repetition suppression in macaque inferotemporal cortex. *J. Neurosci.* **38**, 7492–7504 (2018).
30. J. G. Kim *et al*., Functions of ventral visual cortex after bilateral medial temporal lobe damage. *Prog. Neurobiol.* **191**, 101819 (2020).
31. J. J. Sakon, W. A. Suzuki, A neural signature of pattern separation in the monkey hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9634–9643 (2019).
32. N. A. Suthana *et al*., Specific responses of human hippocampal neurons are associated with better memory. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10503–10508 (2015).
33. M. J. Jutras, P. Fries, E. A. Buffalo, Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13144–13149 (2013).
34. J. D. Dworkin *et al*., The extent and drivers of gender imbalance in neuroscience reference lists. *bioRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.01.03.894378.
35. D. Maliniak, R. Powers, B. F. Walter, The gender citation gap in international relations. *Int. Organ.* **67**, 889–922 (2013).
36. N. Caplar, S. Tacchella, S. Birrer, Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nat. Astron.* **1**, 0141 (2017).
37. D. Zhou *et al*., Gender diversity statement and code notebook v1. 0 *Zenodo* (2020).