Multiplicative mixing of object identity and image attributes in single inferior temporal neurons

N. Apurva Ratan Murty[a,1] and S. P. Arun[a,2]

[a]Centre for Neuroscience, Indian Institute of Science, 560012 Bangalore, India

Edited by Thomas D. Albright, The Salk Institute for Biological Studies, La Jolla, CA, and approved February 27, 2018 (received for review August 14, 2017)

Object recognition is challenging because the same object can produce vastly different images, mixing signals related to its identity with signals due to its image attributes, such as size, position, rotation, etc. Previous studies have shown that both signals are present in high-level visual areas, but precisely how they are combined has remained unclear. One possibility is that neurons might encode identity and attribute signals multiplicatively so that each can be efficiently decoded without interference from the other. Here, we show that, in high-level visual cortex, responses of single neurons can be explained better as a product rather than a sum of tuning for object identity and tuning for image attributes. This subtle effect in single neurons produced substantially better population decoding of object identity and image attributes in the neural population as a whole. This property was absent both in low-level vision models and in deep neural networks. It was also unique to invariances: when tested with two-part objects, neural responses were explained better as a sum than as a product of part tuning. Taken together, our results indicate that signals requiring separate decoding, such as object identity and image attributes, are combined multiplicatively in IT neurons, whereas signals that require integration (such as parts in an object) are combined additively.

object recognition | invariance | separability | efficient coding | computer vision

Despite tremendous advances in computing, object recognition remains an extremely challenging problem (1–3). This is, in part, because the same object can produce images that vary in size, position, orientation, and depth depending on its location relative to the observer. As a result, the image impinging on the retina contains signals unique to the identity of an object mixed with signals related to its specific image attributes (i.e., its image size, image position, viewpoint, etc.). How does the brain represent object identity and image attributes so as to enable efficient recognition?

Efforts in understanding this question have focused on the inferior temporal (IT) cortex in the monkey, an area critical for object recognition (4–6). Early proposals focused on the existence of highly invariant (grandmother) cells that encode object identity while discarding all image attributes (7). This idea has largely been discredited, because most IT neurons are modulated by both object identity and attributes, such as position and size (8–10). This has recently been reconfirmed by the fact that neural activity in IT can be used to decode object identity across changes in attributes (11–13) as well as the image attributes themselves (13, 14). Recent studies have shown that the same cells are strongly modulated by both object identity and image attributes (15, 16) and that they also maintain their object preference across size, position, and orientation (16). These findings show that both object identity and image attribute signals are encoded independently by IT neurons but do not specify how they might be combined or what would be an efficient way to do so.

We compared two distinct mechanisms by which these signals might be combined: additively or multiplicatively. Although the sum and product are closely related, we reasoned that this subtle difference can have large functional consequences for how the underlying signals can be decoded. For instance, when two signals are added, a small signal might modulate the sum much less than a large signal, making the smaller signal difficult to decode. In contrast, when signals are multiplied, a small signal can modulate the product as effectively as a large signal, allowing both signals to be easily decoded.

We tested these possibilities by recording from IT neurons using natural objects sampled across a variety of viewing conditions. In all cases, neural responses were accurately explained as a product rather than sum of tuning for object identity and tuning for image attributes such as size, position, and viewpoint. By comparing the information available with additively and multiplicatively mixed responses, we found that multiplicative mixing yielded better decoding of both object identity and attributes. This multiplicative mixing was absent both in low-level vision models as well as in deep convolutional neural networks, but it tended to increase across successive layers of these networks. This property was unique to invariances: when tested with objects created by combining parts, neural responses were better explained as a sum (not product) of part tuning.

Results

We compared two specific ways according to which neurons might combine object identity and image attribute signals: adding or multiplying them. To illustrate these possibilities, consider two simulated neurons with identical tuning for objects and image attributes (say size). In the first neuron, the response to a particular object ($o1$) presented at a particular size ($s2$) would be $r(o1,s2)=x_o(o1)x_a(s2)$, where x_o and x_a represent object and size tuning, respectively. Thus, its responses combine object and size tuning multiplicatively (Fig. 1*A*). In the second neuron, its response is given by $r(o1,s2)=x_o(o1)+x_a(s2)$; in other words, its responses combine object and size tuning additively (Fig. 1*B*). It can be seen that multiplying these signals results in more selective responses than when adding them and therefore, leads to

Significance

Vision is a challenging problem because the same object can produce a variety of images on the retina, mixing signals related to its identity with signals related to its viewing attributes, such as size, position, rotation, etc. Precisely how the brain separates these signals to form an efficient representation is unknown. Here, we show that single neurons in high-level visual cortex encode object identity and attribute multiplicatively and that doing so allows for better decoding of each signal.

Author contributions: N.A.R.M. and S.P.A. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Present address: McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139.

[2]To whom correspondence should be addressed. Email: sparun@iisc.ac.in.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714287115/-/DCSupplemental.

Published online March 20, 2018.

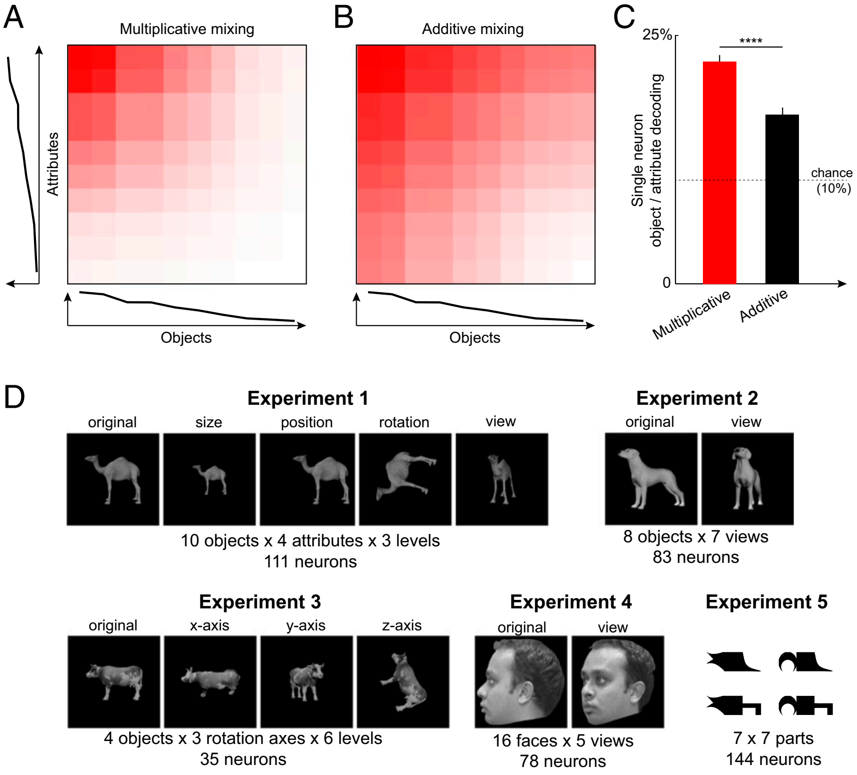

Fig. 1. Multiplicative vs. additive mixing. (*A*) Response of a simulated neuron with identical tuning for objects and attributes (10 levels each) that is combined multiplicatively. It can be seen that multiplying the two signals results in sharply tuned responses. (*B*) Response of a simulated neuron with the same tuning as in *A* but combined additively. It can be seen that adding the signals produces broadly tuned responses. (*C*) Average accuracy of object or attribute decoding for simulated single neurons with randomly initialized but identical tuning for both object and attribute. It can be seen that multiplying signals leads to more accurate decoding than adding them. Error bars represent SEM across iterations. The dashed line represents chance decoding (1 of 10 objects = 10%). Asterisks represent statistical significance. ****$P < 0.00005$, sign rank test on decoding accuracy across 100 simulated neurons. (*D*) Schematic of five experiments in this study. In experiments 1–4, we manipulated objects along several image attributes. Experiment 1 comprised the same set of objects manipulated in size, position, in-plane rotation, or 3D view. Experiment 2 consisted of objects manipulated across several 3D views. Experiment 3 consisted of objects rotated along the cardinal axes. Experiment 4 consisted of faces rotated in depth. In experiment 5, we tested part integration in objects by creating a large number of objects by combining the same seven parts on the left or right side.

better decoding of both signals. To quantify this, we trained a linear classifier on the response of each simulated neuron to decode object identity or attribute (*Methods*). Across many randomly chosen tuning functions, we obtained consistently better decoding from neurons with multiplicative mixing compared with additive mixing (Fig. 1*C*). When the underlying object and size tuning are unknown, these can be estimated by recording the neural response to many objects across many sizes. This allows us to ask whether the complex response properties of IT neurons can be explained by additive or multiplicative mixing.

We hypothesized that IT neurons might combine signals multiplicatively when they require independent decoding but additively when they require signal integration. We tested this hypothesis across five experiments, which are summarized in Fig. 1*D*. In experiments 1–4, we tested IT neurons on images of objects varying in size, position, in-plane rotation, and in-depth rotations. In experiment 1, we investigated neural responses to images of objects varying along a number of identity-preserving attributes: size, position, in-plane rotation, and in-depth rotations about the *y* axis. In experiments 2–4, we investigated in-depth rotations in greater detail for objects (experiment 2), cardinal axis rotations (experiment 3), and faces (experiment 4). Across all experiments, neural responses were explained better as a product (not sum) of tuning for object identity and attributes. In experiment 5, we tested IT neurons using two-part objects. Here, we predicted that part signals will require integration and therefore, combine additively.

Object Tuning Across Multiple Image Attributes (Experiment 1). Here, we recorded the responses of 127 neurons to objects with balanced changes in size, position, orientation (i.e., in-plane rotation), and viewpoint (i.e., rotations in depth). The response of a representative IT neuron is depicted in Fig. 2*A* for all objects across all image attributes. In the resulting color map, strong responses along a column indicate that the neuron prefers a particular object across all attributes, and strong responses along a row indicate preference for a particular attribute (size/position/rotation/view) across all objects. These patterns, in turn, indicate separable tuning for objects and their image attributes. To quantify how combined tuning for object identity and image attribute can be predicted by individual tuning for identity and tuning for attributes, we fit a multiplicative model to the observed responses (Fig. 2*B*). To avoid overfitting, we used one-half of all trials to estimate separate tuning for object identity and attribute signals (*Methods*), generated a predicted response for objects across attributes, and compared it with the observed response on the other one-half of the trials. For this neuron, we obtained an excellent correlation between observed and predicted responses ($r = 0.86$, $P < 0.00005$) (Fig. 2*C*) that was as good as the consistency of its firing across two halves of trials (mean ± SD across many split halves: $r = 0.80 \pm 0.03$). Other examples of observed responses and multiplicative model fits are shown in Fig. 2*D*.

This pattern was true across the entire population as well: multiplicative model predictions were correlated with observed responses (average correlation: $r = 0.55 \pm 0.19$ across 112 neurons with a significant split-half correlation) across a majority of cells (108 of 112 cells showed a significant correlation, $P < 0.05$). This relatively low model correlation could stem from noisy neural firing or from systematic variations in the response that cannot be explained by the model. To assess this possibility, we reasoned that the upper bound for any model derived from odd trials to predict the firing rate on even trials would simply be the degree to which odd trials themselves predict even trials. Therefore, we calculated a normalized correlation for each neuron, wherein we divided the model correlation by the split-half correlation. The normalized correlation for the multiplicative model was close to 1, indicating that it explains all of the explainable variance in the response (Fig. 2*E*) (normalized correlation, mean ± SD: 1.10 ± 0.26 across 112 neurons with a significant split-half correlation; values larger than 1 typically came from neurons with low split-half correlation). This high degree of fit persisted even on assessing each attribute separately (normalized correlation, mean ± SD: 1.07 ± 0.18, 1.05 ± 0.25, 1.02 ± 0.19, and 1.02 ± 0.23 for size, position, rotation, and view, respectively). This high degree of fit was present in both early and late stages of the neural response (normalized correlation, mean ± SD: 0.99 ± 0.44 for firing rates during 0–100 ms; 1.09 ± 0.22

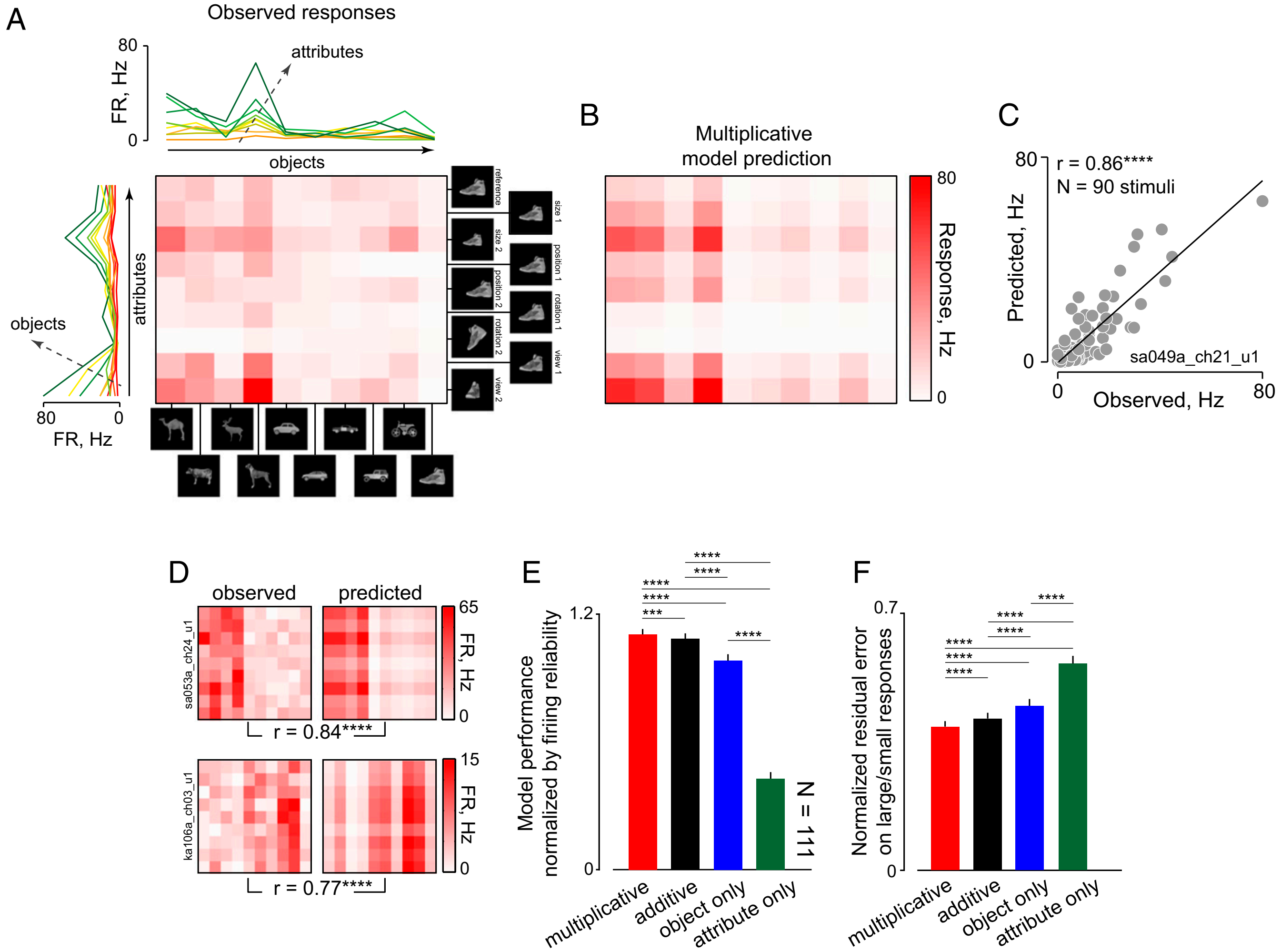

Fig. 2. Objects across many attributes (experiment 1). (*A*, *Lower Right*) Observed responses for an example IT neuron across objects (along rows) and various attributes (along columns). (*A*, *Upper*) Responses to objects for various attributes and (*A*, *Left*) responses to attributes across objects. FR, firing rate. (*B*) Predicted responses for the multiplicative model for this neuron. (*C*) Observed response plotted against predicted response across all 90 stimuli. Asterisks indicate statistical significance, and the solid line is the best-fitting line. $****P < 0.00005$. (*D*) Observed responses and multiplicative model predictions for two other example IT neurons. $****P < 0.00005$. (*E*) Mean and SEM of model correlation (normalized by firing reliability) across neurons for the multiplicative (red), additive (black), object-only (blue), and attribute-only (green) models. Asterisks indicate statistical significance on a Wilcoxon sign rank test comparing pairs of model correlations across neurons. $***P < 0.005$; $****P < 0.00005$. (*F*) Residual error between model predictions and observed firing for stimuli that elicited large/small firing rates (identified by z scoring individual cell responses and selecting stimuli with $|z| > 2$), where additive and multiplicative models are expected to differ in their predictions. $****P < 0.00005$.

for firing rates during 100–200 ms), suggesting that multiplicative separability remains stable over time. Thus, the multiplicative model explained virtually all of the systematic variation in the neural firing.

To be sure that the multiplicative model was indeed the best model, we compared it with several alternative models. The primary alternative was an additive model, in which object and image attribute tuning add instead of multiply. We note that this model is difficult to distinguish from the multiplicative model, because the sum and product of two numbers always covary: for instance, the sum and product of two sets of 100 numbers generated using a Poisson process with mean 10 spikes per 1 s are strongly correlated ($r = 0.98$, $P < 10^{-72}$). Nonetheless, the sum and product of two numbers produce subtly but quantitatively different predictions, using which they can be distinguished, particularly when the numbers are disparate. Indeed, the normalized correlation of the additive model was slightly but significantly worse compared with the multiplicative model (Fig. 2*E*). Likewise, on calculating the residual error of the two models, the multiplicative model produced smaller residual errors for large/small observed firing rates (Fig. 2*F*). Finally, the advantage of the multiplicative model over the additive model was apparent at all levels of firing reliability (*SI Text*).

We also considered an object-only model, which considered only object preferences and discarded all image attribute modulation. This model yielded considerably worse fits to the data, indicating that very few neurons were perfectly invariant (Fig. 2 *E* and *F*). Finally, we considered an attribute-only model, which considered only image attribute preferences and discarded all object identity modulation. This model too yielded considerably worse fits compared with the multiplicative model (Fig. 2 *E* and *F*).

Do Fully Invariant Cells Carry More Information? The above results show that, on average, the response of IT neurons can be explained using a product of identity and image attribute tuning. However, there may be smaller subgroups of invariant cells that may be the neural substrate for invariant object recognition. To evaluate this possibility, we performed an ANOVA on the firing rate across trials of each neuron with object identity and attribute as factors. A majority of all neurons (83 of 127 or 65%) were modulated by both identity and attribute, whereas only 18% of cells (23 of 127) were modulated only by object identity but not by attribute (i.e., potentially invariant). Are the invariant cells more selective (and therefore, more informative) for objects compared with the cells with both identity and attribute effects? To assess this possibility, we measured the sparseness of each neuron across objects (*Methods*) and compared the two groups of cells. Cells modulated by both identity and attribute were more sharply tuned for objects compared with the cells modulated only by identity (average object sparseness: 0.06 for 23 object-only cells; 0.12 for 83 object and attribute cells, $P < 0.005$, rank sum test on sparseness). Likewise, cells modulated by both identity and attribute yielded more accurate decoding of object identity than cells modulated only by identity (decoding accuracy: 43% for cells with identity and attribute modulation, 22% for object-only cells, $P < 0.00005$ using bootstrap sampling; chance performance = 10%). Thus, cells with multiplicative modulation of identity and attribute are the majority of cells in IT cortex and are strongly selective for object identity.

The above results are based on testing neurons using object images that varied along four attributes: size, position, orientation, and viewpoint. In doing so, we equated objects to have the same overall size, position, etc. before systematically changing these attributes (we note that equating viewpoint is nontrivial; see below). To confirm that equating attributes across objects was indeed important, we shuffled the responses of each object so that objects are no longer equated for their attributes. Shuffled responses were fit equally well by the object-only model as by the multiplicative model (normalized correlation, mean ± SD: 0.96 ± 0.27 for multiplicative model, 0.98 ± 0.32 for the object-only model, $P = 0.24$, sign rank test across 112 cells). Thus, equating object attributes was critical to establish the multiplicative separability of identity and attributes.

Does Multiplicative Separability Lead to More Invariant Population Decoding? So far, we have found a relatively small but significant advantage of the multiplicative model over the additive model in single neurons. The simulation in Fig. 1 already shows that even a subtle difference in a single neuron can be functionally relevant by enabling better decoding. However, this simulation was based on taking arbitrary tuning functions and combining them. We, therefore, sought to confirm whether this decoding advantage for multiplicative mixing would hold given the range of stimulus selectivity observed in IT neurons.

To investigate this issue, we created two groups of simulated neurons derived from the observed neural responses as depicted in Fig. 3*A*. In the first group, we took the multiplicative model prediction for each observed neuron and generated noisy firing rates for eight trials (same as in the observed data), each using a Poisson process. In the second group, we took the additive model predictions for each neuron and then generated noisy firing rates using a Poisson process. We then compared the ability of linear classifiers to decode object identity from the additive and multiplicative neural populations. We found that the multiplicative population had a decoding accuracy that was substantially higher than the additive population (Fig. 3*B*) (average accuracy for object decoding: 96 and 87% for the simulated multiplicative and additive populations, respectively; additive accuracy was never larger than multiplicative accuracy across 1,000 bootstrap samples obtained by repeatedly sampling 50 randomly chosen cells; thus, $P < 0.001$). This was true for image attribute decoding as well (average accuracy: 65 and 52% for multiplicative and additive populations, respectively; additive accuracy exceeded multiplicative accuracy in 6 of 1,000 bootstrap samples; thus, $P = 0.006$). Thus, multiplicative separability at the single-neuron level leads to improved decoding of both object identity and image attribute at the level of the entire population.

Do Results Generalize to Other Objects and Image Attributes (Experiments 2–4)? The results of experiment 1 were based on testing objects across image attributes, such as size, position, orientation, and viewpoint. Of these, size, position, and orientation are straightforward image transformations, since they leave image features fundamentally unchanged. As a result, it was simple to equate objects for their size, position, or orientation. However, viewpoint changes are qualitatively different: when an object is rotated in depth, its features can appear, disappear, compress, or expand. This makes it nontrivial to equate objects across changes in viewpoint. This, in turn, raises the possibility that the multiplicative separability observed in experiment 1 was driven largely by changes in size, position, and orientation and not by changes in viewpoint. We therefore systematically investigated this issue in experiments 2–4.

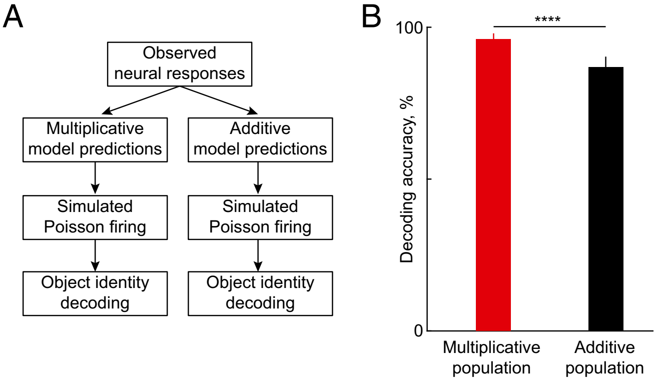

Fig. 3. Decoding with multiplicative and additive neurons. (*A*) Although multiplicative and additive model performance was only subtly different, it could have significant implications for population decoding. To investigate this issue, we created two populations of neurons from the observed neural data in experiment 1. In the first group of neurons, we took additive model predictions, simulated Poisson firing for the same number of trials as in the original experiment, and trained linear classifiers to decode object identity. In the second group, we performed the same analysis using multiplicative model predictions. (*B*) Decoding accuracy for the simulated multiplicative and additive populations (chance = 10%). Asterisks indicate statistical significance calculated as the fraction of bootstrap samples (by sampling 50 neurons each time), in which the decoding accuracy was larger for the additive population. ****$P < 0.0005$.

In experiment 2, we tested 113 IT neurons on objects across rotations about the *y* axis. We chose *y*-axis rotations as a starting point, because they are the most frequently encountered in natural vision (for other axis rotations, see below). To equate objects across viewpoint, we designed objects to all have a single impoverished view and sampled viewpoints on either side (Fig. 4). The responses of an example IT neuron are illustrated in Fig. 4*A*. This neuron responded strongly to the sideways profile view of all objects and least of all to the impoverished view, which may be expected, since very few image features are visible at the impoverished view. Importantly, its response was predicted extremely well by the multiplicative model (Fig. 4*B*) with a strong correlation ($r = 0.90$, $P < 0.0005$) (Fig. 4*C*). This model fit was close to the split-half reliability of its response (mean ± SD of split-half correlation: 0.85 ± 0.04).

The observed responses and the multiplicative model predictions for two other IT neurons are shown in Fig. 4*D*. It can be seen that the multiplicative model produces excellent fits to the neural responses. This was true in general across neurons (Fig. 4*E*) (normalized correlation, mean ± SD: 0.97 ± 0.25 across 83 cells with a significant split-half correlation). Thus, the multiplicative model explained nearly all of the systematic variation in neural firing. As before, we compared the performance of the multiplicative model across neurons with the performance of an additive model, an object-only model, and a view-only model. The performance of the multiplicative model was significantly better than the other models both in terms of normalized correlation (Fig. 4*E*) as well as using residual error calculated on large/small firing rates (Fig. 4*F*).

Next, we compared the invariant cells (i.e., only object identity effects) with cells that showed both identity and view effects. As

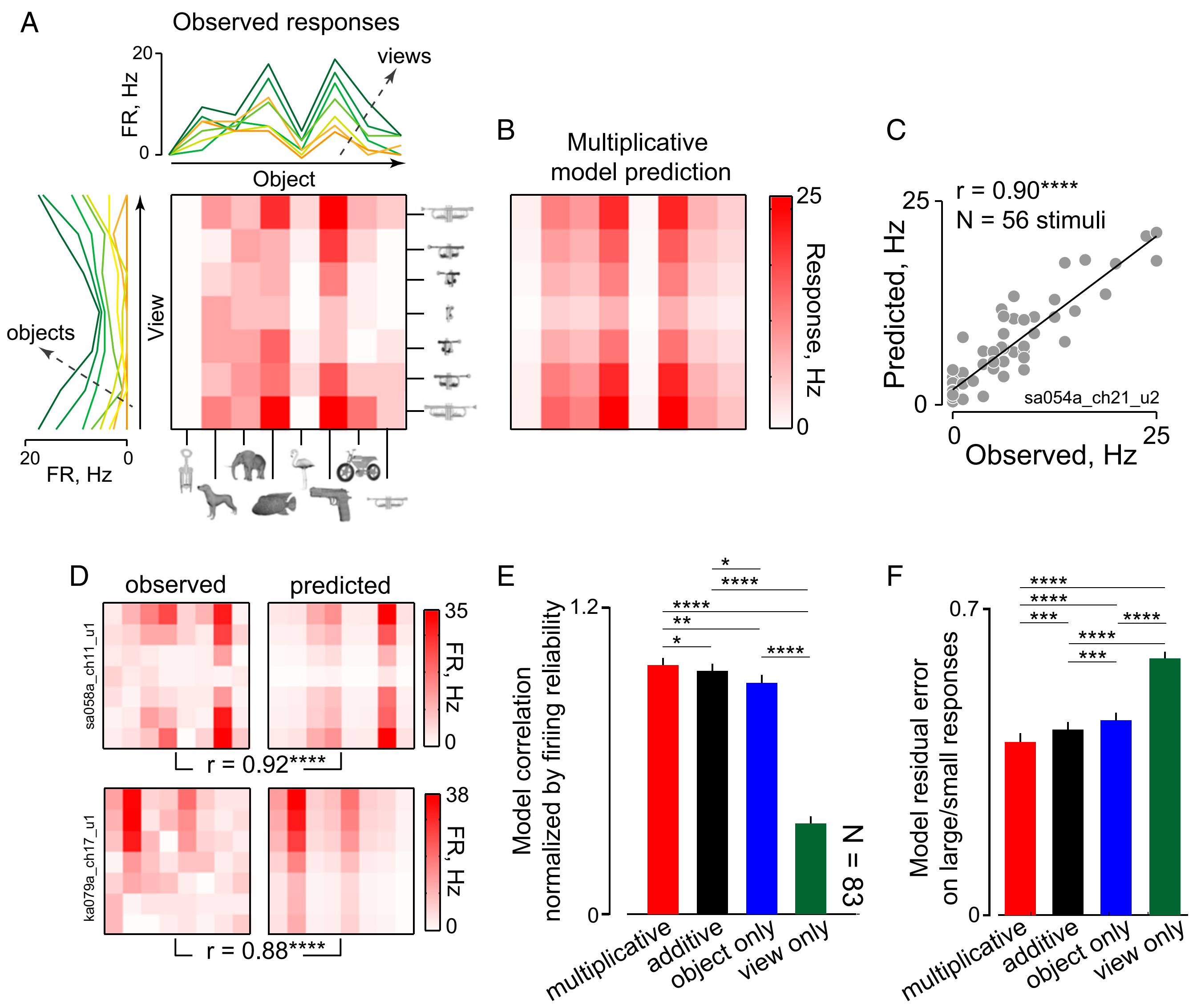

Fig. 4. Objects across rotations in depth (experiment 2). (*A*) Observed responses for an example IT neuron for a set of objects (along columns) presented at many views (along rows). FR, firing rate. (*B*) Predicted responses for the multiplicative model for this neuron. (*C*) Observed response plotted against the multiplicative model prediction across all 56 stimuli, with conventions as before. ****$P < 0.00005$. (*D*) Observed responses and multiplicative model predictions for two other IT neurons. ****$P < 0.00005$. (*E*) Mean and SEM of model correlation (normalized by firing reliability) across neurons for the multiplicative (red), additive (black), object-only (blue), and attribute-only (green) models. Asterisks indicate statistical significance on a Wilcoxon sign rank test comparing pairs of model correlations across neurons. *$P < 0.05$; **$P < 0.005$; ****$P < 0.00005$. (*F*) Residual error between model predictions and observed firing for stimuli that elicited large/small firing rates (identified by *z* scoring individual cell responses and selecting stimuli with $|z| > 2$), where additive and multiplicative models are expected to differ in their predictions. Asterisks represent statistical significance. ***$P < 0.0005$; ****$P < 0.00005$.

before, cells with both effects were the greatest in number (52 of 113; i.e., 46%) and were more selective for objects compared with the invariant cells (average sparseness: 0.16 across 52 cells with object and view effects, 0.07 across 32 cells with object-only effects; $P < 0.0005$, rank sum test on sparseness values across cells). Likewise, object identity could be decoded better from cells with object and attribute modulation compared with cells with object-only effects (decoding accuracy: 54% for cells with both effects, 49% for object-only cells, chance = 12.5%, P = 0.01 using bootstrap sampling). Thus, cells modulated by both identity and image attribute convey more information about object identity than the invariant cells with object-only effects.

Can Objects Really Be Equated Across Viewpoints? Our results so far show that, for objects rotated about a single impoverished view, IT neurons show multiplicative separability for identity and viewpoint. However, can objects be equated for viewpoint at all? To investigate the effect of object structure further, we recorded the responses of the same neurons to objects with two impoverished views instead of only one. These objects had qualitatively different viewpoint relations as measured using population neural dissimilarity (*SI Text*). This, in turn, implies that including these objects will lead to a breakdown of multiplicative separability. Indeed, as predicted, neural responses were multiplicatively separable for objects with consistent viewpoint relations but not for objects with inconsistent viewpoint relations (*SI Text*).

In experiments 1 and 2, we investigated objects across rotations about the *y* axis. In experiment 3, we investigated objects across rotations about all three cardinal axes (*x*, *y*, and *z*). Once again, neural responses were multiplicatively separable (*SI Text*). In experiment 4, we tested IT neurons for faces across many views. We selected faces as a special case, where equating objects

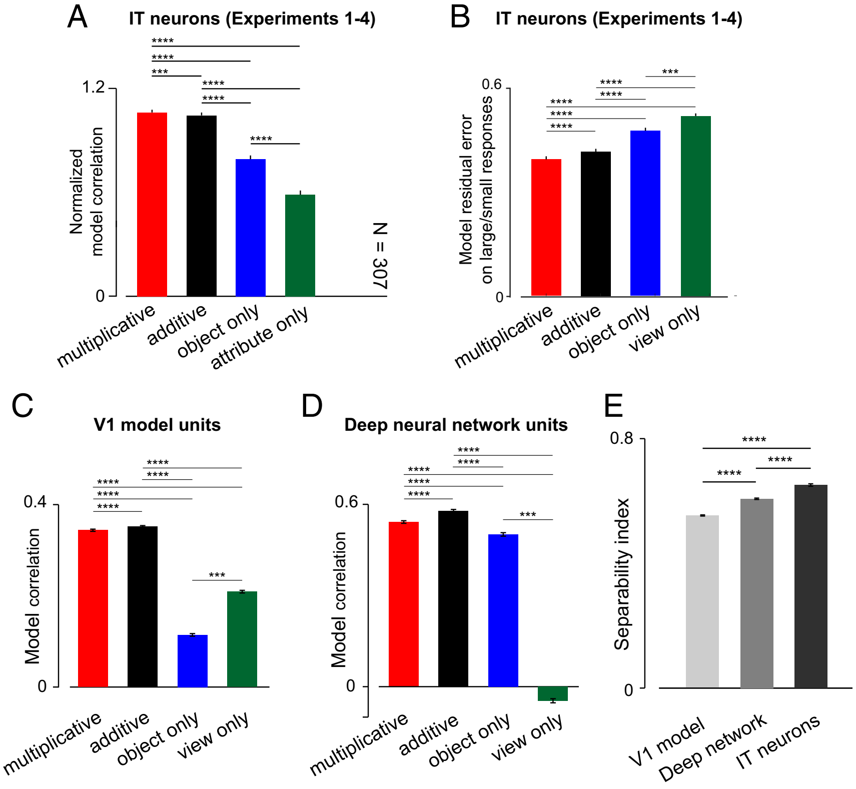

Fig. 5. Separability in IT neurons and computational models (experiments 1–4). (*A*) Normalized model correlation for the multiplicative (red), additive (black), object-only (blue), and attribute-only (green) models across all recorded neurons across all experiments. Error bars indicate the SEM across neurons. Asterisks indicate statistical significance as before based on a Wilcoxon sign rank test comparing performance across neurons for each pair of models. (*B*) Residual error between model predictions and observed firing for stimuli that elicited large/small firing rates (identified by *z* scoring individual cell responses and selecting stimuli with $|z| > 2$), where additive and multiplicative models are expected to differ the most in their predictions. Asterisks represent statistical significance as before based on a sign rank test on comparing average error across neurons. (*C*) Same as *B* but for the V1 model units. (*D*) Same as *B* but for deep neural network units. (*E*) Magnitude of multiplicative separability measured using the separability index for the V1 model, the deep neural network model, and IT neurons. Error bars indicate the SEM across all units. Asterisks indicate statistical significance based on a Wilcoxon rank sum test comparing separability indices across models. $^{***}P < 0.0005$; $^{****}P < 0.00005$.

across views is straightforward. Here too, the multiplicative model yielded better fits to the data (*SI Text*).

Model Performance Across Experiments 1–4. To summarize, we have tested IT neurons on four diverse object sets: objects varying along multiple attributes (experiment 1), objects across *y*-axis rotations (experiment 2), objects across all cardinal axis rotations (experiment 3), and faces across viewpoint changes (experiment 4). We combined the model performance for the multiplicative model, additive model, object-only model, and attribute-only model across all experiments to obtain a global summary of our findings and compare with computational models. The multiplicative model yielded consistently outperformed all models both in terms of normalized correlation (Fig. 5*A*) as well as in terms of residual error (Fig. 5*B*). We conclude that object identity and image attributes are multiplicatively separable at the level of IT neurons.

Heterogeneity of Signal Mixing in Single Neurons. So far, we have compared the aggregate behavior of the multiplicative and additive models, but there could be considerable variability across single neurons. We investigated this possibility in several ways. First, we sought to compare the multiplicative and additive models for each neuron to ascertain the numbers of neurons that favored each model. To do so, we compared the residual error across stimuli for each neuron. The residual error of the two models was significantly different ($P < 0.05$, sign rank test) in only 46 of 307 cells across experiments 1–4. This relatively small number of cells detected is not surprising given the tight correlation between sums and products in general as well as the limited numbers of trials per stimulus. However, 82% (38 of 46) of these cells had a smaller residual error for the multiplicative model, and this fraction was significantly different from the 50:50 split expected by chance ($P < 0.00005$, χ^2 test). Thus, while there are individual cells that favor additive mixing, such cells are relatively few in number and are outnumbered by cells that favor multiplicative mixing.

Second, we considered the possibility that individual neurons might implement a broad continuum of signal integration ranging from additive to multiplicative mixing of identity and attribute signals. To investigate this possibility, we fit a mixed model, in which the response R is given by $R = a{*}R_a + m{*}R_m$, where R_a and R_m are the additive and multiplicative predictions and a and m are scalars representing their contributions. A purely multiplicative response would have $a = 0$, whereas a purely additive response would have $m = 0$. On fitting this model, the multiplicative term was significantly larger than the additive one in experiments 1–4 (*SI Text*). Furthermore, the mixed model yielded fits that were better than the additive model but no better than the multiplicative model. Thus, while there is variation across cells in the extent of additive vs. multiplicative signal mixing, the aggregate tendency in the IT population favors multiplicative mixing of object identity and image attribute signals.

Multiplicative Separability in Computational Models. The above findings show multiplicative separability of object identity and attributes in IT neurons for identity-preserving attributes but not identity-altering transformations. However, this could be trivially inherited from low-level visual areas, or alternatively, it could be an emergent property in high-level visual cortex. To address this issue, we tested two computational models on the images used in experiments 1–4. The first model was a V1 model (13, 17). If V1 model units show multiplicative separability, then it is likely to be inherited by downstream visual areas. The second model was a deep convolutional neural network optimized for object classification (18). Such deep networks have been extremely successful in predicting response properties of neurons along the ventral stream (19–21). Of particular interest to us was whether deep neural network units would show an increasing multiplicative separability, which would indicate that this is a computational requirement for invariant object recognition.

We present the combined performance of all models here for simplicity (individual experiments are in *SI Text*). Without fitting these models to the IT data, we analyzed individual units in these models exactly as we did with the IT data. As before, we fit multiplicative, additive, object-only, and attribute-only models to the responses of each model unit and concatenated model performance on stimulus sets across all four experiments. With IT neurons, we had responses across many trials for each image, and therefore, we were able to use the more robust split-half cross-validation procedure. However, since model activations do not vary across trials, we used leave-one-out cross-validation to evaluate model performance. As a result, while model performance in absolute terms cannot be directly compared with IT neurons, it was possible to evaluate whether the response of each unit can be explained best using additive, multiplicative, object-only, or attribute-only models.

For the V1 model, we found that individual unit activations were explained best using an additive model (Fig. 5*C*) and not a multiplicative model. Thus, low-level visual representations show additive rather than multiplicative separability. We found a similar result with the deep neural network. Individual units responses to objects across attributes were better explained using an additive model rather than the multiplicative model (Fig. 5*D*). However, both the additive and multiplicative models performed increasingly better across layers (*SI Text*), suggesting that overall separability (regardless of type) is an emergent property across layers in the network.

How does the multiplicative separability observed in IT neurons compare with that observed in computational models? To assess this possibility, we calculated an index of multiplicative separability on the full response of each neuron (*Methods*) that represents the fraction of the overall variance in the response that is accounted for by the multiplicative model. Across experiments, multiplicative separability was largest for IT neurons followed by deep network units and smallest for V1 model units (Fig. 5*E*). Thus, IT neurons have the most efficient representation in terms of multiplicative separability. Interestingly, the deep neural network had greater multiplicative separability compared with the V1 representation, but it was still smaller than the separability in IT. We propose that separability in general and multiplicative separability in particular are desirable properties for an invariant object representation.

Do Parts in an Object Also Combine Multiplicatively (Experiment 5)?

In experiments 1–4, we have shown that neural responses to objects across varying attributes can be explained using a product but not sum of tuning for objects and attributes. In experiment 5, we asked whether such multiplicative separability would occur for objects with discrete parts. Our motivation was that part signals are more likely to be integrated rather than being constrained for independent decoding, like in the case of objects and attributes. We therefore surmised that part signals might combine additively rather than multiplicatively.

To investigate these issues, we recorded the responses of 180 IT neurons to objects created by combining two parts on either end of a stem in a combinatorial manner (*Methods*). The responses of an example IT neuron (using firing rates in a 50- to 250-ms window) to the full set of objects are shown in Fig. 6*A*. It can be seen that the neuron responds strongly to all objects sharing a particular part. The predictions of the additive model (Fig. 6*B*) were as strongly correlated with the observed response (r = 0.6) (Fig. 6*C*) as the reliability of firing itself (split-half correlation, mean ± SD; r = 0.57 ± 0.06). Thus, the additive model captured nearly all of the systematic variation in neural firing. The observed and predicted responses of two other example IT neurons are shown in Fig. 6*D*.

This pattern was true across the neural population: the additive model explained nearly all of the systematic variation in firing as evidenced by a highly normalized correlation (Fig. 6*E*). Interestingly, the additive model outperformed all other models both in terms of overall match to the data (Fig. 6*E*) as well as in terms of residual error for large/small firing rates (Fig. 6*F*). Thus, neural responses to objects with discrete parts are explained as a sum—not product—of part signals.

Discussion

Here, we have shown that object identity and attribute signals combine multiplicatively in IT neurons across diverse objects varying along many attributes. This multiplicative separability was efficient in that it led to better object decoding than additive separability. It was nontrivial in that low-level and deep neural network representations show additive rather than multiplicative separability. It may be an emergent property for invariant recognition, since separability increases across layers of deep neural networks optimized for object recognition. Finally, multiplicative separability did not always occur: part signals within an object combine additively in IT neurons rather than multiplicatively. Together, these findings suggest that signals that require separate decoding might combine multiplicatively in IT neurons and that signals that require integration combine additively. Below, we review our findings in the context of the existing literature.

Our finding that identity and attribute signals combine multiplicatively is consistent with frequent observation that IT neurons maintain their object preference across size and position (8, 9, 22) and across viewpoint (13, 22). Our finding that part signals combine additively in IT neurons is consistent with similar observations made for discrete parts (23) and shape/color (24). It is also consistent with the fact that IT neurons preserve their shape tuning across textures (25, 26). It does not agree, however, with the finding that face features combine multiplicatively in face cells (27)—whether this is specific to faces or face cells remains an interesting open question. Except for this last study, most others have not explicitly compared multiplicative and additive mixing as we have done. Distinguishing these two possibilities is nontrivial, because the sum and product of two numbers are always strongly correlated and can be distinguished primarily when the two numbers are disparate. Even in our study, multiplicative and additive models were only slightly but significantly different in their predictions. However, we have additionally shown that these subtle differences at the single-neuron level lead to substantial differences in decoding object identity or attribute for the neural population as a whole (Fig. 3).

Our observation of multiplicative separability depended critically on equating attributes across objects. This was straightforward for size, position, and orientation but nontrivial for in-depth rotations. For viewpoint, we found multiplicative separability only when objects have consistent viewpoint relations, such as when they are aligned to an impoverished view, but not when they have inconsistent viewpoint relations (*SI Text*). This is concordant with the idea that objects may undergo viewpoint transitions as they are rotated in depth (28). However, we have explicitly evaluated viewpoint relations using neural dissimilarity rather than using abstract shape features. We propose that evaluating neural or perceptual dissimilarity between views can be a powerful approach to studying viewpoint relations between and across objects.

Our finding that object/attribute signals combine multiplicatively but part/part signals combine additively raises several interesting questions. First, why might this occur? We have shown that multiplying signals allows for efficient decoding of either signals. This might be useful for decoding object identity and attribute separately. However, additive mixing might be more useful when object parts have to be integrated to represent whole objects (24). Alternatively, it could be that object attributes, being irrelevant variations, are combined multiplicatively, whereas parts, being relevant variations, are combined additively. Distinguishing between these possibilities will require training animals on categorizing stimuli with relevant and irrelevant features. Second, why is there heterogeneity at the single-cell level? While we have observed that the average tendency across neurons favors multiplicative mixing, there was considerable heterogeneity in signal mixing (*SI Text*). It is possible that there are neurons that perform additive mixing or even more complex tuning functions with activity that becomes relevant in

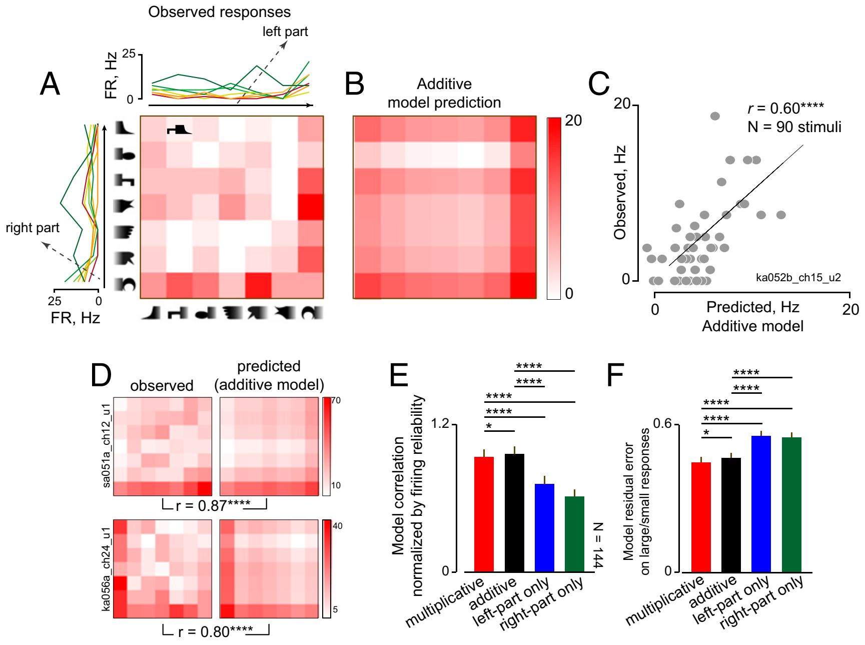

Fig. 6. Objects with varying parts (experiment 5). (*A*) Observed responses for an example IT neuron for a set of objects with varying right parts (along columns) and left parts (along rows). Individual parts corresponding to each row and column are depicted but were never shown in the experiment. An example object is depicted in the first row, second column. FR, firing rate. (*B*) Predicted responses for the additive model for this neuron. (*C*) Observed response plotted against the additive model prediction across all 49 stimuli, with conventions as before. *****P* < 0.00005. (*D*) Observed responses and multiplicative model predictions for two other IT neurons. *****P* < 0.00005. (*E*) Mean and SEM of model correlation (normalized by firing reliability) across neurons for the multiplicative (red), additive (black), object-only (blue), and attribute-only (green) models. Asterisks indicate statistical significance on a Wilcoxon sign rank test comparing pairs of model correlations across neurons. **P* < 0.05; *****P* < 0.00005. (*F*) Residual error between model predictions and observed firing for stimuli that elicited large/small firing rates (identified by *z* scoring individual cell responses and selecting stimuli with |*z*| > 2), where additive and multiplicative models are expected to differ in their predictions. Asterisks represent statistical significance as in *E*.

specialized tasks. Testing this will require evaluating neural responses across different task contexts. Third, what are the underlying mechanisms? Both additive and multiplicative mixing can be accomplished using the neural mechanism of divisive normalization that is prevalent throughout visual cortex (29–31). Specifically, it has been shown in the context of attentional modulation that divisive normalization can result in a broad range of response modulations from additive to multiplicative (31).

Multiplicative separability is a common motif in many brain regions. It is best known in gain fields in parietal cortex (32), but it has been observed in auditory cortex (33) as well as in multiple visual areas for disparate features, such as motion/disparity (34) and orientation/disparity (35). While it is well-established that IT neurons are invariant to size, position, and viewpoint, the fact that object identity and attributes combine multiplicatively represents a unique finding. We have also shown that multiplicative separability is nontrivial in that it is absent in low-level visual representations and increases along successive layers of deep neural networks optimized for object classification. These findings show that multiplicative separability is an emergent property of neural networks optimized for object recognition. More generally, we propose that multiplicative separability emerges in the brain whenever multiple signals need to be combined while allowing for efficient decoding of either.

Methods

All animal experiments were performed according to a protocol approved by the Institutional Animal Ethics Committee of the Indian Institute of Science, Bangalore and the Committee for the Purpose of Control and Supervision of Experiments of Animals, Government of India. Most experimental procedures are similar to those reported in previous studies from our laboratory (36) and are, therefore, only briefly summarized below.

Neurophysiology. We recorded from the left IT cortex of two macaque monkeys (*Macacca radiata*; Ka and Sa, age 7 y old) using standard neurophysiological procedures detailed previously (36). Recording sites were verified using MRI to be in the anterior ventral portion of the IT cortex. Extracellular wideband signals were recorded at 40 KHz using 24-channel laminar electrodes (Uprobe; 100-μm intercontact spacing; Plexon Inc.) linked to a neural data acquisition system (Plexon Inc.). These signals were manually sorted offline into distinct clusters using spike sorting software (OfflineSorter; Plexon Inc.). Only well-isolated visually responsive units were selected for further analyses. The numbers of recorded neurons in each experiment are reported below.

Behavioral Task. Each animal was trained to fixate a series of stimuli presented at the center of gaze. Each trial began after the animal fixated on a small red fixation dot (0.2°), after which eight stimuli were presented for 200 ms each with an interstimulus duration of 200 ms. Images within a trial were presented in random order with the constraint that no two images of an object occurred one after the other to avoid response adaptation. Error trials were repeated after a random number of other trials. Each stimulus was repeated about 8–11 times across trials. Monkeys received a juice reward at the end of each trial for successfully maintaining fixation throughout the trial.

Experiment 1: Objects Across Multiple Attributes. The stimuli in this experiment comprised 10 objects (5 animate, 5 inanimate) with images that were systematically varied in size, position, orientation, and viewpoint. Objects were equated across attributes by scaling, shifting, and rotating a reference image to have the same size, position, and orientation. All objects were chosen such that they had an impoverished view, at which most of their features were obscured, and a most elongated view, in which most of their features were visible. Each attribute had three levels, including the reference image. Thus, there were a total of nine unique images corresponding to each object, bringing the total number of stimuli to 90. This dataset has been reported in a recent study (37), but the analyses reported here are unique to this study. In all, we recorded the responses of 127 visually responsive neurons across two monkeys (83 from Ka, 44 from Sa), but for the analyses reported here, we selected a subset of 111 neurons with reliable firing (*P* < 0.05 for the correlation between firing rates estimated from even and odd trials).

Experiment 2: Objects Across *y*-Axis Rotations. The stimuli consisted of 10 objects (4 animate, 4 inanimate, 2 view-inconsistent objects), each presented in seven views. Of these, eight objects had an impoverished view, at which they were the least elongated in the horizontal direction and at which most of their features were obscured. We selected seven viewpoints for each object corresponding to rotations of ±60°, ±30°, ±15°, and 0° about the *y* axis relative to the impoverished view. The remaining two objects had two impoverished views that were at ±30° relative to the other objects. All stimuli were rendered using a 3D modeling software (Autodesk 3DS Max). We recorded from a total of 113 visual neurons in this experiment (49 from Ka, 64 from Sa), but for the analyses reported here, we selected a subset of 83 neurons with reliable firing ($P < 0.05$ for the correlation between firing rates estimated from even and odd trials).

Experiment 3: Objects Across Cardinal Axis Rotations. The stimuli comprised four objects rotated by several levels about each of the three cardinal axes. All objects were equated to have roughly the same 3D volume (and consequently, view relations). From the reference left profile view of each object, we rendered 60°, 120°, 180°, 240°, and 300° rotations about the *x*, *y*, and *z* axes using a 3D modeling software (Autodesk 3Ds Max). We recorded from a total of 50 IT neurons in this experiment (42 from Ka, 8 from Sa), but for the analyses reported here, we selected a subset of 34 neurons with reliable firing ($P < 0.05$ for the correlation between firing rates estimated from even and odd trials).

Experiment 4: Faces Across Rotations in Depth. There were 160 stimuli in this experiment. The first 80 stimuli consisted of the face–object–body subset, which included 10 human faces, 10 animal faces, 20 objects, 10 human bodies, 10 animal bodies, 10 human body parts, and 10 monkey body parts. These stimuli were used to determine neural selectivity for faces, objects, and body parts (related analyses are in *SI Text*). The remaining 80 stimuli consisted of 16 human faces (6 females) photographed in five views corresponding to rotations of ±90°, ±45°, and 0° about the front-facing view. We recorded from 117 neurons in this experiment (97 from Ka, 20 from Sa), but for the analyses reported here, we selected a subset of 78 neurons with reliable firing ($P < 0.05$ for the correlation between firing rates estimated from even and odd trials).

Experiment 5: Objects Created by Combining Parts. There were 49 stimuli in this experiment. Each stimulus was an object created by adding two distinct parts on either side of a horizontal stem. The full stimulus set was created by combining seven possible parts on the left and right sides in all possible ways. We recorded from 180 neurons in this experiment (93 from Ka, 87 from Sa), but for the analyses reported here, we selected a subset of 144 neurons with more reliable firing ($P < 0.5$ for the correlation between firing rates estimated from even and odd trials). This dataset has been reported previously (38), but the analyses reported here are unique to this study.

Data Analysis.

Single-neuron analysis of decoding (Fig. 1C). To compare multiplicative and additive mixing in terms of their ability to decode either property, we took single neurons with random (but identical) tuning functions for object identity and image attribute as shown in Fig. 1*A* and created neural responses that were either multiplicative (Fig. 1*A*) or additive (Fig. 1*B*). We then trained a linear classifier to decode object identity or attribute in a leave-one-out fashion. The accuracy of this classifier represents the degree to which object identity (or attribute) could be decoded given the response of a single noiseless neuron.

Model fitting. For each neuron, we calculated its firing rate during the image presentation period (0–200 ms) in odd-numbered trials and created a response matrix **R** with entries r_{ij} representing the response to the *j*th object at the *i*th attribute. For example, in experiment 1 (with 10 objects and 9 attributes), the response matrix **R** has 10 columns and 9 rows (as shown in Fig. 2*A*). We then fit this response matrix to five possible models as detailed below.

i) Additive model. According to the additive model, the neural response can be written as $r_{ij} = a_i + o_j$, where a_i is the unknown activation due to attribute *i* and o_j is the unknown activation due to object *j*. For example, in experiment 1, this would imply 10 unknown activations for objects and 9 unknowns for attributes, resulting in a total of 19 unknowns. To estimate these activations, we averaged the response matrix along the rows to obtain the attribute activations $[a_1, a_2, \ldots a_m]$ and along the columns to obtain the object activations $[o_1, o_2, \ldots o_n]$. We then calculated model predictions using the equation $r_{ij} = k_1 a_i + k_2 o_j$, where k_1 and k_2 are constants estimated using linear regression. Note that both attribute and object activations need not be organized along any continuous dimensions as long as attributes are equated across objects.

ii) Multiplicative model. According to the multiplicative model, the neural response can be written as $r_{ij} = a_i o_j$, where a_i is the unknown activation due to attribute *i* and o_j is the unknown activation due to object *j*. Thus, the multiplicative model has the same number of free parameters as the additive model. To estimate these unknown activations, we note that the response matrix **R** can be written as an outer product of two vectors $\boldsymbol{R} = \boldsymbol{x_a} \boldsymbol{x_o^T}$, where $\boldsymbol{x_a}$ is the vector $[a_1, a_2, \ldots a_m]$ containing the attribute activations and $\boldsymbol{x_o}$ is the vector $[o_1, o_2, \ldots o_n]$ containing object activations. Following previous studies (33, 34), we estimated these two activation vectors using singular value decomposition (SVD). This method factorizes the response matrix **R** as $\boldsymbol{R} = \boldsymbol{U \Sigma V^T}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are matrices containing the left and right singular vectors, respectively, and $\boldsymbol{\Sigma}$ is a diagonal matrix containing the singular values. The multiplicative model output is calculated as the product of the first singular value with the outer product of the first left and right singular vectors. In other words, the multiplicative model prediction is given by $\boldsymbol{R} = \boldsymbol{u}_1 s_1 \boldsymbol{v}_1^T$, where $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are the first column vectors of the $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices, respectively, and s_1 is the first entry of the diagonal matrix $\boldsymbol{\Sigma}$. To be absolutely sure that the superior fits of the multiplicative model over the additive model were not due to the SVD method, we calculated the multiplicative model predictions by multiplying the row and column averages of the response matrix **R**. Here too, the product yielded significantly better fits to the data compared with the sum (normalized correlation, mean ± SEM: 1.10 ± 0.27 for the product, 1.08 ± 0.27 for the sum, $P <$ 0.00005, sign rank test across 111 neurons in experiment 1). We obtained qualitatively similar results for other experiments.

iii) Object-only model. According to the object-only model, the neural response is driven solely by object identity, with no modulation from attributes, and is, therefore, perfectly invariant. The predictions of this model were obtained from the response matrix **R** by averaging along the attribute dimension.

iv) Attribute-only model. According to this model, the neural response is driven solely by image attribute, with no modulation from object identity. The predictions of this model were obtained from the response matrix **R** by averaging along the object dimension.

v) Mixed model. According to this model, the neural response is driven by a mix of additive and multiplicative signals. Specifically, for each stimulus, the response $R = a{*}R_a + m{*}R_m$, where R_a and R_m are the additive and multiplicative predictions for that stimulus, respectively, and *a* and *m* are scalars representing their contributions. We obtained the best-fitting values of *a* and *m* through linear regression.

Model validation. Having obtained model predictions from neural responses on odd-numbered trials, we obtained a cross-validated measure of performance by calculating the Pearson's correlation between these predictions with the firing rate on even-numbered trials. To estimate an upper bound on model performance, we calculated the "split-half" correlation between the firing rates estimated from odd and even trials. We then obtained a normalized measure of model performance by dividing model correlation by the split-half correlation. A normalized correlation close to one indicates that the model explains nearly all of the explainable variance in the response.

Model fitting for computational models. The above approach of training on odd trials and testing on even trials could not be used for computational models, because they produce identical responses with no trial variability. Therefore, we used a leave-one-out cross-validation procedure: we set aside the response to one stimulus each time and calculated the predicted response from all four models for this left-out response. In this manner, we compiled the response to all stimuli and then compared it with the observed response by calculating the correlation coefficient. We fit each model as before, but for the multiplicative model, it was not possible to perform SVD, since the response matrix was no longer complete. We, therefore, multiplied the row and column averages to obtain the multiplicative model predictions.

Multiplicative separability index. To compare multiplicative separability in computational models with that observed in IT neurons, we took the full neural response to objects across attributes and performed an SVD. The ratio of the first singular value to the sum of all singular values represents the fraction of the total variance explained by the first multiplicative outer product, and we took this to indicate the degree of multiplicative separability of the neural response. Specifically, the multiplicative separability index (*SI*) is given as

$$SI = \frac{s_1}{\sum_{i=1}^{n} s_i},$$

where $s_1, s_2, \ldots s_n$ are the singular values from the SVD of the response matrix **R**. This index has a maximum value of one, which indicates complete separability.

Sparseness. To calculate a measure of object selectivity for each neuron, we used a standard measure of sparseness (15, 16). For a neuron with responses $r_1, r_2, \ldots r_n$ to n stimuli, the sparseness is given by $S=(1-(\sum r_i/n)^2/(\sum r_i^2/n))/(1-1/n)$, where the summation is across all responses. For an extremely sparse neuron that responds to only one stimulus in a set, the sparseness is one. For a broadly tuned neuron that responds equally to all stimuli, the sparseness is zero. Thus, a large value of sparseness indicates a more selective response. We calculated object selectivity for each neuron by taking its average response to objects (across attributes) and calculating sparseness. We obtained similar results using other measures of tuning.

Population decoding. To characterize the nature of information available in the neural population, we performed a population decoding analysis on single-trial neural responses. We took the firing rate of each neuron evoked during a 50- to 200-ms window after image onset as the value along each dimension of a multidimensional vector space. We trained a linear classifier on these response vectors corresponding to individual trials of each stimulus, with the class labels being either object or attribute. Note that this approach assumes that responses were recorded simultaneously, but this provides an upper bound on the information available to the entire population if recorded simultaneously. To measure the information conveyed about invariant object identity, we trained a classifier on responses to objects at one size and tested it on the responses at another size. This was done for all pairs of attributes to obtain an average decoding estimate.

ACKNOWLEDGMENTS. We thank Carl Olson for insightful comments; Zhivago, Pramod, Puneeth, and Harish for being an excellent recording team; and V. Ramesh of the Primate Research Laboratory for outstanding animal care. This research was funded by a Council for Scientific and Industrial Research Senior Research Fellowship (to N.A.R.M.) and an Intermediate Fellowship from the Wellcome Trust–Department of Biotechnology India Alliance and the Department of Biotechnology–Indian Institute of Science Partnership Programme (to S.P.A.).

1. DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73:415–434.
2. Everingham M, et al. (2014) The pascal visual object classes challenge: A retrospective. *Int J Comput Vis* 111:98–136.
3. Pramod RT, Arun SP (2016) Do computational models differ systematically from human object perception? *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ), pp 1601–1609.
4. Logothetis NK, Sheinberg DL (1996) Visual object recognition. *Annu Rev Neurosci* 19: 577–621.
5. Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139.
6. Connor CE, Brincat SL, Pasupathy A (2007) Transformation of shape information in the ventral pathway. *Curr Opin Neurobiol* 17:140–147.
7. Gross CG (2002) Genealogy of the "grandmother cell." *Neuroscientist* 8:512–518.
8. Ito M, Tamura H, Fujita I, Tanaka K (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73:218–226.
9. Brincat SL, Connor CE (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7:880–886.
10. Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189.
11. Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310:863–866.
12. Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol* 97:4296–4309.
13. Ratan Murty NA, Arun SP (2015) Dynamics of 3D view invariance in monkey inferotemporal cortex. *J Neurophysiol* 113:2180–2194.
14. Hong H, Yamins DLK, Majaj NJ, DiCarlo JJ (2016) Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat Neurosci* 19:613–622.
15. Zoccolan D, Kouh M, Poggio T, DiCarlo JJ (2007) Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci* 27:12292–12307.
16. Zhivago KA, Arun SP (2016) Selective IT neurons are selective along many dimensions. *J Neurophysiol* 115:1512–1520.
17. Pinto N, Cox DD, DiCarlo JJ (2008) Why is real-world visual object recognition hard? *PLoS Comput Biol* 4:e27.
18. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556v6.
19. Güçlü U, van Gerven MA (2015) Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J Neurosci* 35:10005–10014.
20. Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLOS Comput Biol* 10:e1003915.
21. Yamins DLK, et al. (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci USA* 111:8619–8624.
22. Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE (2008) A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci* 11:1352–1360.
23. Sripati AP, Olson CR (2010) Responses to compound objects in monkey inferotemporal cortex: The whole is equal to the sum of the discrete parts. *J Neurosci* 30:7948–7960.
24. McMahon DBT, Olson CR (2009) Linearly additive shape and color signals in monkey inferotemporal cortex. *J Neurophysiol* 101:1867–1875.
25. Arcizet F, Jouffrais C, Girard P (2009) Coding of shape from shading in area V4 of the macaque monkey. *BMC Neurosci* 10:140.
26. Köteles K, De Mazière PA, Van Hulle M, Orban GA, Vogels R (2008) Coding of images of materials by macaque inferior temporal cortical neurons. *Eur J Neurosci* 27:466–482.
27. Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12:1187–1196.
28. Tarr MJ, Kriegman DJ (2001) What defines a view? *Vision Res* 41:1981–2004.
29. Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61: 168–185.
30. Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9: 181–197.
31. Ghose GM, Maunsell JH (2008) Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. *J Neurosci* 28:5115–5126.
32. Salinas E, Thier P (2000) Gain modulation: A major computational principle of the central nervous system. *Neuron* 27:15–21.
33. Peña JL, Konishi M (2001) Auditory spatial receptive fields created by multiplication. *Science* 292:249–252.
34. Smolyanskaya A, Ruff DA, Born RT (2013) Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. *J Neurophysiol* 110:2806–2816.
35. Grunewald A, Skoumbourdis EK (2004) The integration of multiple stimulus features by V1 neurons. *J Neurosci* 24:9185–9194.
36. Ratan Murty NA, Arun SP (2017) Seeing a straight line on a curved surface: Decoupling of patterns from surfaces by single IT neurons. *J Neurophysiol* 117:104–116.
37. Ratan Murty NA, Arun SP (2017) A balanced comparison of object invariances in monkey IT neurons. *eNeuro* 4:1–10.
38. Pramod RT, Arun SP (2018) Symmetric objects become special in perception because of generic computations in neurons. *Psychol Sci* 29:95–109.