Visual attention mitigates information loss in small- and large-scale neural codes

Thomas C Sprague[1,*], **Sameer Saproo**[2], and **John T Serences**[1,3,*]

[1]Neurosciences Graduate Program, University of California San Diego, La Jolla, CA, USA 92093-0109

[2]Department of Biomedical Engineering, Columbia University, New York, New York, USA

[3]Department of Psychology, University of California San Diego, La Jolla, CA, USA 92093-0109

Summary

The visual system transforms complex inputs into robust and parsimonious neural codes that efficiently guide behavior. Because neural communication is stochastic, the amount of encoded visual information necessarily decreases with each synapse. This constraint requires processing sensory signals in a manner that protects information about relevant stimuli from degradation. Such selective processing – or selective attention – is implemented via several mechanisms, including neural gain and changes in tuning properties. However, examining each of these effects in isolation obscures their joint impact on the fidelity of stimulus feature representations by large-scale population codes. Instead, large-scale activity patterns can be used to reconstruct representations of relevant and irrelevant stimuli, providing a holistic understanding about how neuron-level modulations collectively impact stimulus encoding.

Keywords

Vision; visual attention; stimulus reconstruction; information theory; neural coding

Visual attention and information processing in visual cortex

Complex visual scenes contain a massive amount of information. To support fast and accurate processing, behaviorally relevant information should be prioritized over behaviorally irrelevant information. For example, when approaching a busy intersection while driving it is critical to detect changes in your lane's traffic light rather than one nearby in order to prevent a dangerous collision. This capacity for selective information processing, or *selective visual attention*, is supported by enhancing the amount of information that is encoded about relevant visual stimuli relative to the amount of information that is encoded about irrelevant stimuli. Importantly, understanding how relevant visual stimuli are

---

[*]Correspondence: tsprague@ucsd.edu (TCS). jserences@ucsd.edu (JTS).

represented with higher fidelity requires considering more than just the impact of attention on the response properties of individual neurons. Instead, examining activity patterns across large neural populations can provide insights into how different unit-level attentional modulations synergistically improve the quality of stimulus representations in visual cortex.

In the scenario above, neurons can undergo several types of modulation in response to the relevant light compared to when it is irrelevant: response amplitudes can increase (response gain), responses can become more reliable, and receptive field properties can shift (e.g. some neurons will shift their spatial receptive field to encompass the attended light). Thus, neural responses associated with attended stimuli generally have a higher signal-to-noise ratio and are more robust compared to responses evoked by unattended stimuli. Accordingly, the behavioral effects associated with visual attention are thought to reflect these relative changes in neural activity: when stimuli are attended, participants exhibit decreased response times, increased discrimination accuracy, and improved spatial acuity ([1-3] for reviews).

This selective prioritization of relevant over irrelevant stimuli follows from two related principles of information theory [4-7] (Box 1). First, the *data processing inequality* [7] states that information is inevitably lost when sent via noisy communication channels, and that lost information cannot be recaptured via any amount of further processing. Second, the *channel capacity* of a communication system is determined by the amount of information that can be transmitted and received, and by the degree to which that information is corrupted during the process of transmission. In the brain, channel capacity is finite because there is a fixed (albeit large) number of neurons and because synaptic connections are stochastic so information cannot be transmitted with perfect fidelity. Given this framework, different types of attention-related neural modulations can be viewed as a concerted effort to attenuate the unavoidable decay of behaviorally relevant information as it is passed through subsequent stages of visual processing [8,9]. This framing also highlights the importance of understanding how attention differentially impacts responses across neurons, and more importantly, how these modulations at the single-unit level interact to support population codes that are more robust to the information processing limits intrinsic to the architecture of the visual system.

With this goal in mind, we first provide a selective overview of recent studies that examine attentional modulations of single measurement units (e.g. single neurons or single fMRI voxels) in visual cortex, with a focus on changes in response amplitude and shifts in spatial sensitivity profiles. We then introduce a framework for evaluating how attention-induced changes in large-scale patterns of activity can shape information processing to counteract the inherent limits of stochastic communication systems. This approach emphasizes reconstructing representations of sensory information based on multivariate patterns of neural signals and relating the properties of these reconstructions to changes in behavioral performance across task demands.

Attention changes response properties of tuned neurons

Single-neuron firing rates in macaque primary visual cortex (V1) [10-12], extrastriate visual areas V2 and V4 [12-19], motion-sensitive middle temporal area (MT) [20-22], lateral intraparietal cortex (LIP) [23-27], and frontal eye fields (FEF) [28-31] have all been shown to reliably increase when either a spatial position or a feature of interest is attended (Fig. 2A). Heightened neural activity can facilitate the propagation of responses to downstream areas, leading to successively weaker distracter-associated responses compared to target-associated responses. However, even though most studies focus on increases in mean firing rates, many studies also report that a substantial minority of cells show systematic decreases in firing rates with attention (particularly in excitatory cells, e.g. [32]), an important issue when considering population-level neural codes that we revisit below.

In addition to measuring attentional modulations in response to a fixed stimulus set, researchers have also parametrically varied stimuli while an animal maintains a constant focus of attention in order to measure changes in feature tuning functions (TFs) or spatial receptive fields (RFs; that is, a neuron's response profile to each member of a set of stimuli, see Glossary). When an animal is cued to attend to a visual feature, like orientation [15,16], color [33] or motion direction [34], neurons tuned to the cued feature tend to respond more, while those tuned to uncued features tend to respond less [35]. This selective combination of response gain and suppression results in a larger range of possible firing rates, or a larger dynamic range (see Glossary), and thus increases encoding capacity such that different features will evoke a more easily separable neural response (Fig. 2B) [36,37]. This increase in encoding capacity with an increase in dynamic range is analogous to switching between a binary and a grayscale image (e.g., a barcode and a black-and-white photograph): the number of states each pixel can take increases, meaning more states are discriminable.

When attending to a particular spatial position, the spatial RF of many neurons can also shift to accommodate the attended position in V4 [38], MT [39-42], and LIP [43], and the endpoint of a saccadic eye movement in V4 [44] and FEF [45]. In MT, for example, RFs shrink around the locus of attention when animals are cued to attend to a small region within a neuron's spatial RF (Fig. 2C) [13,39,40]. However, when attention is focused just outside the penumbra of a neuron's spatial RF, the RF shifts and expands towards the focus of attention [41]. Finally, the tuning of V4 neurons to orientation and spatial frequency (that is, their spectral receptive field) can undergo shifts towards an attended target stimulus when an animal is viewing natural images [46]. These changes in the size and position of spatial RFs – coupled with increases in response amplitude – may lead to a more robust population code via an increase in the number of cells that respond to relevant features ([2,47,48] for a review). For example, an increase in single-cell firing rates coupled with a shift in the selectivity profile of surrounding spatial RFs towards the locus of attention should generally increase the overall entropy (see Glossary) of a population code, and thereby the quality of information encoded about a relevant stimulus (Box 1).

Attentional modulation of large-scale populations

Thus far, we have discussed attentional modulations measured from single neurons in behaving monkeys. However, perception and behavior are thought to more directly depend on the quality of large-scale population codes [36,49,50], so it is also necessary to assay how these small-scale modulations jointly impact the information content of larger-scale neural responses. For example, human neuroimaging methods like fMRI and EEG provide a window into the activity of large-scale neural populations [51-53], enabling the assessment of attention-related changes in voxel- or electrode-level signals that reflect the aggregate responses of all constituent neurons.

The firing rate increases observed in single neurons are echoed by attention-related increases in fMRI BOLD activation levels [54-59] and amplitude increases in stimulus-evoked EEG signals [60-65]. For example, when attention is directed to one of several stimuli on the screen, the mean BOLD signal measured from visual cortical regions of interest (ROIs) increases [55-59,66,67]. Additionally, when fMRI voxels are sorted based on their selectivity for specific features such as orientation [37,68,69] (see also [70]), color [71], face identity [72], or spatial position [73], attention has the largest impact on voxels that are tuned to the attended feature value.

In addition to changes in response amplitude, recently developed techniques can also assess changes in the selectivity of voxel-level tuning functions across different attention conditions. One newly developed method has been used to evaluate how the size of voxel-level population receptive fields (pRFs) changes with attentional demands [74,75]. For example, several studies have measured pRF size as participants view a display consisting of a central fixation point and a peripheral visual stimulus that is used to map the pRF. On different trials, participants either attend to the peripheral mapping stimulus, or they ignore the mapping stimulus and they instead attend to the central fixation point. Attending to the peripheral mapping stimulus increases the average size of voxel-level pRFs measured from areas of extrastriate cortex where single-neuron RFs are relatively large. However, no such size modulations are observed in primary visual cortex where single-neuron RFs are smaller [76-78]. At first, this result appears to conflict with neurophysiology studies showing that single-neuron RFs can either shrink or expand depending on the spatial relationship between the neuron's RF and the focus of attention (see above: Fig 2C, [39-42]). However, the response of a voxel reflects the collective response of all single neurons that are contained in that voxel. As a result, when the attended mapping stimulus was anywhere in the general neighborhood of the voxel's spatial RF, many single-neuron RFs within the voxel likely shifted towards the attended stimulus. In turn, this shifting of single-neuron RFs towards attended stimuli in the vicinity of the voxel's RF should increase the area of visual space over which the voxel would respond (i.e. it would increase the size of the pRF compared to when the fixation point is attended and these neuron-level shifts would not occur).

The above studies examined how voxel-level pRFs change when the RF mapping stimulus is attended. A complementary line of work has addressed how attention to a focused region of space alters pRFs measured using an unattended mapping stimulus. In these studies, attention was directed either to the left or the right of fixation while a visual mapping

stimulus was presented across the full visual field. The center of voxel-level pRFs shift towards the locus of attention [79]; however, because the authors do not report whether pRFs also change in size, it is challenging to fully interpret how shifting the center of a pRF would support enhanced encoding of attended information (Box 1, Fig. 3). Furthermore, another study found that increasing the difficulty of a shape discrimination task at fixation leads to a shift of voxel-level pRFs away from fixation and also to an increase in their size. These modulations may thus result in a lower-fidelity representation of irrelevant stimuli in the visual periphery when a foveated stimulus is challenging to discriminate [80].

Functionally similar examples of information shunting have also been found in other domains: Brouwer and Heeger [71] demonstrated that directing attention to a colored stimulus during a color categorization task narrows the bandwidth of voxel-level tuning functions, improving the discriminability of voxel responses for distinct colors when the color value was important for the task. Similarly, Çukur et al [81] used a high-dimensional encoding model (which describes the visual stimulus categories for which each voxel is most selective) to show that attending to object categories shifts semantic space towards the attended target to increase the number of voxels responsive to a relevant category (see also [82,83]). While analogous single-unit neural data are not available for comparison, these results support the notion that shifting the feature selectivity profile of a RF is an important strategy implemented by the visual system to combat limited channel capacity via increasing the sampling density for relevant information (Fig. 1).

With all experiments evaluating responses at the large-scale population level (e.g., the level of a single fMRI voxel), it is important to note that these macroscopic measurements reflect hemodynamic signals related to net changes in the response across hundreds of thousands or more neurons [51]. As a result, it is currently not possible to unambiguously infer whether attentional modulations measured at the scale of single-voxels reflect changes in neuron-level feature selectivity or if voxel-level modulations instead reflect non-uniform changes in the response amplitude of neural populations within a voxel that are tuned to different feature values. In spite of this limitation, large-scale measurement techniques like fMRI can provide a unique perspective on the collective impact of small-scale single-neuron modulations on the fidelity of population codes, even though information about the specific pattern(s) of single-neuron modulations may be obscured. In turn, changes in voxel-level selectivity can support some important general inferences about the impact of attention on the encoding capacity of large-scale population responses (Box 1), which are not easily accessible via single-neuron recording methods.

Reconstructing region-level stimulus representations

The techniques used to measure single-neuron and single-voxel response profiles help us understand how changes at the level of single measurement units (whether single neurons or single voxels) can impact encoding capacity to facilitate perception and behavior. But understanding how individual encoding units behave, either in isolation or at the level of a population average, is only a part of the picture. Indeed, different neurons and voxels are often modulated in different ways even in the context of the same experimental design: some units increase their response amplitude, others decrease [32]; some show (p)RF size

increases, others show decreases [39-42,76-78,80]. To understand how these apparently disparate modulations work together to impact the quality of region-level population codes, multivariate methods can be used to directly infer changes in the overall information content of neural response patterns.

An emerging means of evaluating information content of population-codes is via *stimulus reconstruction* (Box 2). Although population-level reconstruction methods have existed for decades [84,85], they have recently found widespread application in the field of human neuroimaging [71,75,86-94]. There are many variations on these methods, but all generally involve first estimating an *encoding model* that describes the selectivity profile (feature TF or spatial RF) of individual measurement units (e.g., single neurons or single voxels). Next, these encoding models are inverted and used to reconstruct the stimulus given a novel pattern of responses across the entire set of units (Box 2). Each computed stimulus reconstruction contains a representation of the stimulus of interest. Thus, for features like color, orientation, or motion, the results of this procedure reflect a reconstruction of the response across a set of feature-selective populations [9,71,88-90,95-99]; for models based on spatial position, the results reflect reconstructed images of the visual scene viewed or remembered by an observer [75,76,86,100,101]. We call this broad framework whereby patterns of encoding models are inverted in order to reconstruct stimulus representations *inverted encoding models* (IEM, see glossary).

The ability to reconstruct an image of visual stimuli based on population-level activation patterns can be used to assess how modulations observed at the level of measurement units are combined to jointly constrain the amount of information encoded about a stimulus. In contrast to multivariate classification analyses that partition brain states into one of a set of discrete groups ([91,102]) or Bayesian approaches that generate an estimate of the most likely stimulus feature value [36,49,50,84,103-105], reconstruction enables the quantification of stimulus representations in their native feature space rather than in signal space (see Glossary). In turn, quantifying representations within these reconstructions supports the ability to evaluate attributes such as the amplitude or the precision of the encoded representation. Moreover, because the IEM reconstruction method involves an analog mapping from an idiosyncratic signal space that differs across individuals and ROIs into a common feature space, it is possible to directly compare quantified properties of stimulus reconstructions as a function of attentional demands. This approach thus complements previous efforts to establish the presence of stimulus-specific information by decoding which stimulus of a set was most likely to have caused an observed pattern of activation [75,87,106,107].

Because region-level stimulus reconstructions exploit information contained in the pattern of responses across all measurement units, they may be more closely linked to behavioral measures than responses of single neurons or even mean response changes across a small sample of neurons or brain regions [91,97,108-110]. Additionally, representations within these region-level reconstructions can be subjected to similar information theoretic analyses as described in Box 1 [9]. Instead of comparing how the response of a small sample of neurons changes with attention (Fig. 3A), it is possible to evaluate how all co-occurring

response modulations constrain the ability of a neural population to encode relevant information about a stimulus (Fig. 3B).

Although this approach can provide a unique perspective on the quality of large-scale population codes, stimulus reconstruction methods come at the cost of simplifying assumptions about how information is encoded. For instance, IEMs for simple features (Box 2) will not account for information that is not explicitly modeled. Thus, an IEM for reconstructing spatial representations of simple stimuli [76,101] will not recover any information that was represented about features such as color or orientation, despite the known roles that many visual areas play in encoding these stimulus attributes.

Reconstructions as an assay of population-level information

Attention has been shown to induce a heterogeneous set of modulations at the level of single measurement units such as single cells or voxels. Variability in the magnitude or sign of attention effects is often treated as noise, and the impact of different types of attentional modulation on the quality of stimulus representations is usually not considered (e.g., the joint influence of both gain and bandwidth modulation). IEMs can be used to extend these unit-level results and to evaluate how all types of attentional modulation collectively influence the information content of large-scale population codes.

Similar to the information content of single-unit responses, when the amplitude of a population-level stimulus reconstruction increases above baseline, then more of the variability in the reconstruction is directly linked with changes in the stimulus (i.e. there is an increase in signal entropy). Importantly, in contrast to single units (Fig 3A, see also Box 1), when a population-level stimulus reconstruction becomes more precise, the population may support more precise inferences about stimulus features by improving the discriminability of responses associated with different stimulus feature values (Fig 3B, e.g., [71,96,97]). Such a change in stimulus reconstructions could be supported by changes in the selectivity of individual voxels/neurons, non-uniform application of neural gain across the population, or any combination of these response modulations at the unit-level.

In one study where participants categorized colors, voxel-level tuning functions for hue narrowed, and region-level reconstructions of color response profiles were more clustered in a neural color space compared to when color was irrelevant [71]. This result provides evidence for a neural coding scheme whereby relevant category boundaries for a given task are maximally separated. Since more of the variability in the population response should be associated with changes in the relevant stimulus dimension (greater signal entropy), this modulatory pattern should provide a more robust population code that can better discriminate different exemplars in color space (Fig. 3B, see also [111,112]).

Similarly, directing attention to the direction of a moving stimulus increased the amplitude of direction-selective representations in both V1 and MT relative to attending stimulus contrast (Fig. 2D) [9]. This increase in the dynamic range of responses gives rise to an increase in the information content of the direction-selective representation in both areas (via an increase in signal entropy) when motion was relevant compared to when it was irrelevant. In addition, when attention was directed to motion, the efficacy of feature-

selective information transfer between V1 and MT increased relative to when stimulus contrast was attended (Box 1; Fig. 2F). This task-dependent increase in the transfer of information between brain regions suggests that attention not only modulates the quality of signals within individual cortical regions, but also increases the efficiency with which representations in one region influence representations in another [9]. Although the precise mechanism for such information transfer remains unknown, changes in synaptic efficacy [11] or synchrony of population-level responses such as LFP and/or spike timing [113-115] likely contribute.

Spatial attention can change the amplitude of single-neuron responses and their spatial selectivity (Fig. 2A-C). One study examined how all these changes jointly modulate representations of a visual stimulus within spatial reconstructions of the scene [76]. Participants were asked to perform either a demanding spatial attention task or a demanding fixation task in the scanner. Their fMRI activation patterns were then used to reconstruct stimulus representations from several visual ROIs. Though individual voxel-level pRFs were found to increase in size with attention, no changes were found in the size of stimulus reconstructions with attention. This pattern of results indicates that attention does not sharpen the region-level stimulus representations in this task. However, the study did reveal attention-related increases in the amplitude of stimulus representations (Fig. 2E), which corresponds to more information about the represented stimulus (greater signal entropy) above a noisy, uniform baseline (noise entropy, Box 1; Fig. 3B). These results were echoed by a recent report that linked attention-related increases in the size and the gain of spatial pRFs in ventral temporal cortex with improved population-level information about stimulus position [77].

Concluding remarks and future directions

Selective attention induces heterogeneous modulations across single encoding units, and understanding how these modulations interact is necessary to fully characterize their impact on the fidelity of information coding and behavior. At present, analysis techniques that exploit population-level modulations have primarily been implemented with data from large-scale neuroimaging tools, such as fMRI and EEG. Applying these analyses to other methods such as 2-photon *in vivo* calcium imaging of neurons identified genetically [116,117] or by cortical depth [118] and electrophysiological recordings from large-scale electrode arrays in behaving animals and humans[119,120] will help bridge gaps in our understanding of how the entire range of neuron-level attentional modulations are related to population-level changes in the quality of stimulus representations (Box 3). Furthermore, the development of improved modeling, decoding, and reconstruction methods as applied to both human neuroimaging and animal physiology and imaging data should enable new inferences about the mechanisms of attention in more complicated naturalistic settings, potentially even during unrestrained movement [50,91-94,121].

Additionally, associating neural modulations with changes in behavioral performance is critically important as a gold-standard method for evaluating the impact of attention on the quality of perceptual representations. The importance of this brain-behavior link was recently highlighted by a study in which visual attention was correlated with the modulation

of single-neuron activity in visual cortex (increased firing rate, among others). However, these modulations in visual cortex were unaffected even after attention-related improvements in behavior were abolished by the transient inactivation of the superior colliculus, an area that is thought to play an important role in attentional control [122,123]. This observation places an important constraint on how we consider different mechanisms of attention: attention results in changes to neural codes in visual cortex that should improve the information content about relevant stimulus features compared to irrelevant features (e.g. firing rate modulations). However, these attention-related improvements in the quality of local sensory representations will not necessarily be transmitted to downstream areas, and thus may have little or no impact on behavior. Thus, neural responses, both at the neuron- and population-level, need to be systematically evaluated against changes in behavior in order to establish their overall importance in visual information processing [124].

In future work, one promising approach is to selectively lesion or alter the measured data, post-acquisition, by using only units that show particular encoding or response properties to compute stimulus reconstructions (e.g., measurement units with RFs near or far from the attended stimulus; measurement units that either increase or decrease in response amplitude or RF size [76], Fig. 4). Reconstructions computed using only measurement units with modulations most critical for improving the fidelity of the neural code for attended stimuli compared to unattended stimuli should be associated with increases in mutual information between reconstructions and attended stimuli compared to unattended stimuli, and this can be verified by comparing the information content of reconstructions to measures of behavioral performance across attention conditions [97,98,124]. Using a similar approach, a recent study evaluated the necessity of voxel-level attentional gain on population-level information about spatial position by artificially eliminating attention-related gain from their observed pRFs. They found that pRF gain is not necessary to improve position coding; changes in pRF size and position were sufficient [77]. Finally, the information content of region-level stimulus reconstructions computed in one brain region can be compared to the information content of those measured in other brain regions [125] at different points in time to determine how information is transformed across levels of the visual hierarchy. For example, reconstructions in V1 should primarily reflect information about relevant low-level sensory features, whereas downstream areas in the ventral temporal lobe should encode information about more holistic object properties such as the features associated with relevant faces or scenes. Comparing successive reconstructions across multiple brain regions may highlight those features of visual scenes undergoing attentional selection, how they are selected, and what happens to features not selected (Fig. 1).

By emphasizing the link between the information content of reconstructions across multiple stages of processing and measures of behavioral performance, a more complete picture will emerge about how differently tuned encoding units at each stage – and their associated constellation of attention-induced modulations – can give rise to a stable representation that is more closely linked with the overall perceptual state of the observer.

Box 1

Information content of a neural code

Information is related to a reduction in uncertainty [4,5,7]. A code is informative insofar as measurement of one variable (e.g., a single neuron's firing rate) reduces uncertainty about another variable (e.g., feature of a stimulus). The amount of uncertainty in a random variable (e.g., the outcome of a coin toss or the spiking output of a cell) can be quantified by its entropy (see Glossary), which increases with increasing randomness. Mutual information (MI) is a measure of the reduction in uncertainty of one variable after knowing the state of another variable. MI would be zero for independent variables (e.g., two fair coins), whereas MI would be high for two variables that strongly covary.

If a neuron noisily responds at the same level to each feature value, then the MI between the state of the stimulus and the state of the neuron's response is low, because signal entropy (variability associated with changes in the stimulus) is low and noise entropy (variability unrelated to changes in the stimulus) is high (Fig. IA, see Glossary). Instead, if the neuron exhibits a Gaussian-like orientation tuning function (TF; Fig. IB-C), then MI is higher, as more of the variability in the neuron's response is related directly to changes in the state of the stimulus. In this latter case, if the amplitude of the neuron's TF increases while noise remains approximately constant, then the ratio of signal entropy to noise entropy increases, resulting in greater MI between the neuron's response and the stimulus orientation. However, if the orientation TF changes, this could result in either an increase or decrease in the information about the stimulus, and would be contingent upon several factors, such as the original tuning width, noise structure, dimensionality of the stimulus, and responses of other neurons (Fig. IF-H) [37,126-128]. For a widely-tuned neuron, a decrease in tuning width would result in an increase in signal entropy relative to noise entropy, increasing the information content of the neuron about orientation. At the other extreme, for a neuron perfectly tuned for a single stimulus value with noisy baseline responses to other values, a broadening in tuning would result in greater variability associated with stimulus features, and consequently greater information (Fig. IF-H). Thus, an increase in the amplitude of a neural response (under simple noise models) will increase the dynamic range and entropy, whereas a change in tuning width can either increase or decrease the information content of a neural code.

Box 2

Inverted encoding models enable evaluation of aggregate effects of multiple unit-level response changes on quality of a neural code

When firing rates of single units or activation levels of single voxels are measured in response to several different stimulus values (e.g., orientation of a grating or position of a stimulus on the screen), it is possible to fit an encoding model (see Glossary) to the set of measured responses as a function of feature value. Such an encoding model describes how the neuron or voxel responds to different values of a stimulus, and an accurate encoding model will predict how the neuron or voxel would respond to a novel stimulus value. For example, the best-fit encoding model for an orientation-selective unit would be

a circular Gaussian (Fig. IIA), whereas the encoding model for a spatially selective unit would be characterized by a 2-d Gaussian (Fig. IIB). Note that encoding models need not be visual: measuring firing rates of hippocampal neurons in rodents as they forage for food often reveals a particular region of the environment in which the neuron fires – its place field – which could potentially be described by a 2-d Gaussian encoding model for spatial position within the environment

While the process of estimating encoding models for many single units or single voxels across the brain allows for inferences about the manner in which information is measured, computed, and transformed across different stages of processing, the approach remains massively univariate: all encoding models are estimated in isolation, and inferences about neural processing are based on changes in these univariate encoding models in aggregate [76-80]. In contrast, the inverted encoding model approach (IEM; see Glossary) utilizes the pattern of encoding models estimated across an entire brain region (e.g. primary visual cortex) to reconstruct the region-level representation of a stimulus given a measured pattern of activation across the measurement units (Fig. IIC). These approaches, as implemented presently, rest on assumptions of linearity and are only feasible for simple features of the environment (e.g. orientation, color, spatial position) for which encoding properties are relatively well understood. To date, this general approach has been used to accurately reconstruct feature representations from fMRI voxel activation patterns [9,71,76,88,89,94-101,124], patterns of spike rate in rodent hippocampus [84,119], and human EEG signals measured non-invasively at the scalp [90].

Box 3

Outstanding questions

The IEM technique has recently been adapted for use with scalp EEG signals that can provide insights about the relative timing of attentional modulations of stimulus reconstructions with near-millisecond precision [90]. How do these signals measured with scalp electrodes carry information about features like orientation? And what other types of neural signals (such as 2-photon *in vivo* calcium imaging [116]) can be used for image reconstruction via IEMs?

Correlated variability among neurons is an important limiting factor in neural information processing [108,129-133]. How can this correlated variability be incorporated into the IEM approach, and what are the scenarios in which correlated variability helps and hurts the information content of a population code as measured via stimulus reconstructions?

It is possible to compute region-level feature-specific reconstructions across each of the many visual field maps in cortex [134-136]. However, the role(s) of each of these visual field maps in supporting visual perception and behavior remains largely unknown. By comparing how properties of stimulus representations vary across different visual field maps with measures of behavioral performance, in combination with causal manipulations such as TMS, optogenetic excitation or inhibition of subpopulations of

neurons, and electrical microstimulation, the relative contributions of each region's representation to behavioral output can be compared, and accordingly the role(s) of the region in visual behavior may be inferred.

Application of IEMs for tasks requiring precise maintenance of or attention to visual stimulus features such as orientation, color, or motion direction [9,96,97,124] often reveal different results from those requiring attention to or maintenance of spatial positions [76,98,101]. Attending to a feature sharpens or shifts stimulus reconstructions in a manner well-suited for performing the task, whereas attending to a position enhances the amplitude of stimulus reconstructions over baseline. How do the circumstances in which stimulus reconstructions change in their amplitude differ from those in which reconstructions change their precision?

Acknowledgments

Glossary

Bit	unit of entropy (base 2)
Decoder	algorithm whereby a feature or features about a stimulus (orientation, spatial position, stimulus identity, etc) is/are inferred from an observed signal (spike rate, BOLD signal). Typically, the signal is multivariate across many neurons/voxels, but in principle a decoder can use a univariate signal
Dynamic range	the set of response values a measurement unit can take. An increase in a unit's response gain will increase the range of possible response values, which will increase its entropy
Encoding model	a description of how a neuron (or voxel) responds across a set of stimuli (e.g., a spatial receptive field can be a good encoding model for many visual neurons and voxels, see Box 2)
Entropy	a measure of uncertainty in a random process, such as a coin flip or observation of a neuron's spike count. A variable with a single known value will have 0 entropy, whereas a fair coin would have > 0 entropy (1 bit).
Feature space	after reconstruction using the IEM technique, data exist in feature space, with each data point defined by a vector of values corresponding to the activation of a single feature-selective population response (e.g., orientation, spatial position); common across all participants and visual areas

Inverted encoding model (IEM)	when encoding models are estimated across many measurement units, it may be possible to use all encoding models to compute a mapping from signal space into feature space when allows for reconstruction of stimulus representations from multivariate patterns of neural activity across the modeled measurement units (Box 2).
Multivariate	when analyses are multivariate, signal from more than one measured unit is analyzed; utilizing information about the pattern of responses across units rather than simplifying the data pattern by taking a statistic over the units (e.g., mean)
Mutual Information	amount of uncertainty about a variable (e.g., state of the environment) that can be reduced by observation of a state of the random variable (e.g., voxel or neuron's response)
Noise entropy	variability in one signal that is unrelated to changes in another signal
Receptive field (RF)	region of the visual field which, when visually stimulated, results in a response in a measured neuron or voxel (population RF, or pRF)
Tuning function (TF)	response of a neuron or voxel to each of several values of a feature, such as orientation or motion direction
Signal entropy	variability in one signal that is related to changes in another signal
Signal space	data as measured exist in signal space, with a dimension for each measurement unit (fMRI voxel, EEG scalp electrode, electrocorticography subdural surface electrode, animal single-cell firing rate or calcium signal); cannot be directly compared across individual subjects without a potentially suboptimal coregistration transformation

References

1. Carrasco M. Visual attention: the past 25 years. Vision Res. 2011; 51:1484–525. [PubMed: 21549742]

2. Anton-Erxleben K, Carrasco M. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. Nat. Rev. Neurosci. 2013; 14:188–200. [PubMed: 23422910]

3. Gardner JL. A case for human systems neuroscience. Neuroscience. 2014 DOI: 10.1016/j.neuroscience.2014.06.052.

4. Shannon CE. A Mathematical Theory of Communication. Bell Syst. Tech. J. 1948; 27:379–423.

5. Cover, T.; Thomas, J. Elements of information theory. Wiley; 1991.

6. Quian Quiroga R, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. Nat. Rev. Neurosci. 2009; 10:173–85. [PubMed: 19229240]

7. Shannon, CE.; Weaver, WB. The mathematical theory of communication. University of Illinois Press; 1963.

8. Tsotsos JK. Analyzing vision at the complexity level. Behav. Brain Sci. 1990; 13:423–445.

9. Saproo S, Serences JT. Attention Improves Transfer of Motion Information between V1 and MT. J. Neurosci. 2014; 34:3586–3596. [PubMed: 24599458]

10. Herrero JL, et al. Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. Neuron. 2013; 78:729–39. [PubMed: 23719166]

11. Briggs F, et al. Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. Nature. 2013; 499:476–80. [PubMed: 23803766]

12. Buffalo EA, et al. A backward progression of attentional effects in the ventral stream. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:361–5. [PubMed: 20007766]

13. Moran J, Desimone R. Selective attention gates visual processing in the extrastriate cortex. Science (80-.). 1985; 229:782–784.

14. Luck SJ, et al. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J. Neurophysiol. 1997; 77:24–42. [PubMed: 9120566]

15. Motter BC. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J. Neurophysiol. 1993; 70:909–19. [PubMed: 8229178]

16. McAdams CJ, Maunsell JHR. Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4 . J. Neurosci. 1999; 19:431–441. [PubMed: 9870971]

17. McAdams CJ, Maunsell JHR. Attention to both space and feature modulates neuronal responses in macaque area V4. J. Neurophysiol. 2000; 83:1751–1755. [PubMed: 10712494]

18. Cook EP, Maunsell JHR. Attentional Modulation of Behavioral Performance and Neuronal Responses in Middle Temporal and Ventral Intraparietal Areas of Macaque Monkey. J. Neurosci. 2002; 22:1994–2004. [PubMed: 11880530]

19. Steinmetz NA, Moore T. Eye Movement Preparation Modulates Neuronal Responses in Area V4 When Dissociated from Attentional Demands. Neuron. 2014; 83:496–506. [PubMed: 25033188]

20. Treue S, Maunsell JHR. Attentional modulation of visual motion processing in cortical areas MT and MST. Nature. 1996; 382:539–541. [PubMed: 8700227]

21. Treue S, Maunsell JHR. Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. J. Neurosci. 1999; 19:7591–7602. [PubMed: 10460265]

22. Treue S, Martinez-Trujillo JC. Feature-based attention influences motion processing gain in macaque visual cortex. Nature. 1999; 399:575–579. [PubMed: 10376597]

23. Bisley JW, Goldberg ME. Neuronal activity in the lateral intraparietal area and spatial attention. Science (80-.). 2003; 299:81–86.

24. Bisley JW, Goldberg ME. Attention, intention, and priority in the parietal lobe. Annu. Rev. Neurosci. 2010; 33:1–21. [PubMed: 20192813]

25. Gottlieb JP, et al. The representation of visual salience in monkey parietal cortex. Nature. 1998; 391:481–4. [PubMed: 9461214]

26. Gottlieb J. From thought to action: the parietal cortex as a bridge between perception, action, and cognition. Neuron. 2007; 53:9–16. [PubMed: 17196526]

27. Arcizet F, et al. A pure salience response in posterior parietal cortex. Cereb. cortex. 2011; 21:2498–506. [PubMed: 21422270]

28. Thompson KG, et al. Neuronal basis of covert spatial attention in the frontal eye field. J. Neurosci. 2005; 25:9479–87. [PubMed: 16221858]

29. Armstrong KM, et al. Selection and maintenance of spatial information by frontal eye field neurons. J. Neurosci. 2009; 29:15621–9. [PubMed: 20016076]

30. Juan C-H, et al. Dissociation of spatial attention and saccade preparation. Proc. Natl. Acad. Sci. U. S. A. 2004; 101:15541–4. [PubMed: 15489272]

31. Squire RF, et al. Prefrontal contributions to visual selective attention. Annu. Rev. Neurosci. 2013; 36:451–66. [PubMed: 23841841]

32. Mitchell JF, et al. Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron. 2007; 55:131–41. [PubMed: 17610822]

33. Motter BC. Neural correlates of attentive selection for color or luminance in extrastriate area V4. J. Neurosci. 1994; 14:2178–2189. [PubMed: 8158264]

34. Martinez-Trujillo JC, Treue S. Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr. Biol. 2004; 14:744–51. [PubMed: 15120065]

35. Maunsell JHR, Treue S. Feature-based attention in visual cortex. Trends Neurosci. 2006; 29:317–322. [PubMed: 16697058]

36. Butts DA, Goldman MS. Tuning Curves, Neuronal Variability, and Sensory Coding. PLoS Biol. 2006; 4:e92. [PubMed: 16529529]

37. Saproo S, Serences JT. Spatial Attention Improves the Quality of Population Codes in Human Visual Cortex . J. Neurophysiol. 2010; 104:885–895. [PubMed: 20484525]

38. Connor CE, et al. Spatial attention effects in macaque area V4. J. Neurosci. 1997; 17:3201–3214. [PubMed: 9096154]

39. Womelsdorf T, et al. Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nat. Neurosci. 2006; 9:1156–1160. [PubMed: 16906153]

40. Womelsdorf T, et al. Receptive Field Shift and Shrinkage in Macaque Middle Temporal Area through Attentional Gain Modulation. J. Neurosci. 2008; 28:8934–8944. [PubMed: 18768687]

41. Anton-Erxleben K, et al. Attention reshapes center-surround receptive field structure in macaque cortical area MT. Cereb. Cortex. 2009; 19:2466–2478. [PubMed: 19211660]

42. Niebergall R, et al. Expansion of MT Neurons Excitatory Receptive Fields during Covert Attentive Tracking. J. Neurosci. 2011; 31:15499–15510. [PubMed: 22031896]

43. Ben Hamed S, et al. Visual receptive field modulation in the lateral intraparietal area during attentive fixation and free gaze. Cereb. cortex. 2002; 12:234–245. [PubMed: 11839598]

44. Tolias AS, et al. Eye Movements Modulate Visual Receptive Fields of V4 Neurons. Neuron. 2001; 29:757–767. [PubMed: 11301034]

45. Zirnsak M, et al. Visual space is compressed in prefrontal cortex before eye movements. Nature. 2014; 507:504–507. [PubMed: 24670771]

46. David SV, et al. Attention to Stimulus Features Shifts Spectral Tuning of V4 Neurons during Natural Vision. Neuron. 2008; 59:509–521. [PubMed: 18701075]

47. Desimone R, Duncan J. Neural Mechanisms of Selective Visual Attention. Annu. Rev. Neurosci. 1995; 18:193–222. [PubMed: 7605061]

48. Zirnsak M, Moore T. Saccades and shifting receptive fields: anticipating consequences or selecting targets? Trends Cogn. Sci. 2014; 18:621–8.

49. Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. Nat. Neurosci. 2006; 9:690–696. [PubMed: 16617339]

50. Graf ABA, et al. Decoding the activity of neuronal populations in macaque primary visual cortex. Nat. Neurosci. 2011; 14:239–245. [PubMed: 21217762]

51. Logothetis NK. What we can do and what we cannot do with fMRI. Nature. 2008; 453:869–878. [PubMed: 18548064]

52. Lima B, et al. Stimulus-related neuroimaging in task-engaged subjects is best predicted by concurrent spiking. J. Neurosci. 2014; 34:13878–91. [PubMed: 25319685]

53. Lopes da Silva F. EEG and MEG: relevance to neuroscience. Neuron. 2013; 80:1112–28. [PubMed: 24314724]

54. Kastner S, Ungerleider LG. Mechanisms of visual attention in the human cortex. Annu. Rev. Neurosci. 2000; 23:315–341. [PubMed: 10845067]

55. Buracas GT, Boynton GM. The effect of spatial attention on contrast response functions in human visual cortex. J. Neurosci. 2007; 27:93–97. [PubMed: 17202476]

56. Murray SO. The effects of spatial attention in early human visual cortex are stimulus independent. J. Vis. 2008; 8:2.1–11.

57. Gandhi SP, et al. Spatial attention affects brain activity in human primary visual cortex. Proc. Natl. Acad. Sci. 1999; 96:3314–3319. [PubMed: 10077681]

58. Scolari M, et al. Functions of the human frontoparietal attention network: Evidence from neuroimaging. Curr. Opin. Behav. Sci. 2014; 1:32–39.

59. Gouws AD, et al. On the Role of Suppression in Spatial Attention: Evidence from Negative BOLD in Human Subcortical and Cortical Structures. J. Neurosci. 2014; 34:10347–10360. [PubMed: 25080595]

60. Itthipuripat S, et al. Changing the spatial scope of attention alters patterns of neural gain in human cortex. J. Neurosci. 2014; 34:112–23. [PubMed: 24381272]

61. Müller MM, et al. The time course of cortical facilitation during cued shifts of spatial attention. Nat. Neurosci. 1998; 1:631–4. [PubMed: 10196572]

62. Lauritzen TZ, et al. The effects of visuospatial attention measured across visual cortex using source-imaged, steady-state EEG. J. Vis. 2010; 10

63. Kim YJ, et al. Attention induces synchronization-based response gain in steady-state visual evoked potentials. Nat. Neurosci. 2007; 10:117–25. [PubMed: 17173045]

64. Störmer VS, Alvarez GA. Feature-Based Attention Elicits Surround Suppression in Feature Space. Curr. Biol. 2014; 24:1985–8. [PubMed: 25155510]

65. Itthipuripat S, et al. Sensory gain outperforms efficient readout mechanisms in predicting attention-related improvements in behavior. J. Neurosci. 2014; 34:13384–98. [PubMed: 25274817]

66. Pestilli F, et al. Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. Neuron. 2011; 72:832–46. [PubMed: 22153378]

67. Kastner S, et al. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron. 1999; 22:751–761. [PubMed: 10230795]

68. Serences JT, et al. Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. Neuroimage. 2009; 44:223–31. [PubMed: 18721888]

69. Scolari M, Serences JT. Basing Perceptual Decisions on the Most Informative Sensory Neurons. J. Neurophysiol. 2010; 104:2266–2273. [PubMed: 20631209]

70. Warren SG, et al. Featural and temporal attention selectively enhance task-appropriate representations in human primary visual cortex. Nat. Commun. 2014; 5:5643. [PubMed: 25501983]

71. Brouwer GJ, Heeger DJ. Categorical clustering of the neural representation of color. J. Neurosci. 2013; 33:15454–65. [PubMed: 24068814]

72. Gratton C, et al. Attention selectively modifies the representation of individual faces in the human brain. J. Neurosci. 2013; 33:6979–89. [PubMed: 23595755]

73. Tootell RB, et al. The retinotopy of visual spatial attention. Neuron. 1998; 21:1409–1422. [PubMed: 9883733]

74. Dumoulin S, Wandell B. Population receptive field estimates in human visual cortex. Neuroimage. 2008; 39:647–660. [PubMed: 17977024]

75. Kay K, et al. Identifying natural images from human brain activity. Nature. 2008; 452:352–355. [PubMed: 18322462]

76. Sprague TC, Serences JT. Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat. Neurosci. 2013; 16:1879–87. [PubMed: 24212672]

77. Kay KN, et al. Attention reduces spatial uncertainty in human ventral temporal cortex. Curr. Biol. 2015 DOI: 10.1016/j.cub.2014.12.050.

78. Sheremata SL, Silver MA. Hemisphere-Dependent Attentional Modulation of Human Parietal Visual Field Representations. J. Neurosci. 2015; 35:508–517. [PubMed: 25589746]

79. Klein BP, et al. Attraction of position preference by spatial attention throughout human visual cortex. Neuron. 2014; 84:227–37. [PubMed: 25242220]

80. De Haas B, et al. Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. Curr. Biol. 2014; 24:R66–R67. [PubMed: 24456976]

81. Çukur T, et al. Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. 2013; 16:763–70. [PubMed: 23603707]

82. Seidl KN, et al. Neural Evidence for Distracter Suppression during Visual Search in Real-World Scenes. J. Neurosci. 2012; 32:11812–11819. [PubMed: 22915122]

83. Peelen MV, et al. Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature. 2009; 460:94–7. [PubMed: 19506558]

84. Zhang K, et al. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. J. Neurophysiol. 1998; 79:1017–44. [PubMed: 9463459]

85. Stanley GB, et al. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. J. Neurosci. 1999; 19:8036–42. [PubMed: 10479703]

86. Thirion B, et al. Inverse retinotopy: Inferring the visual content of images from brain activation patterns. Neuroimage. 2006; 33:1104–1116. [PubMed: 17029988]

87. Miyawaki Y, et al. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron. 2008; 60:915–929. [PubMed: 19081384]

88. Brouwer G, Heeger D. Decoding and Reconstructing Color from Responses in Human Visual Cortex. J. Neurosci. 2009; 29:13992–14003. [PubMed: 19890009]

89. Brouwer G, Heeger D. Cross-orientation suppression in human visual cortex. J Neurophysiol. 2011; 106:2108–2119. [PubMed: 21775720]

90. Garcia J, et al. Near-Real-Time Feature-Selective Modulations in Human Cortex. Curr. Biol. 2013; 23:515–522. [PubMed: 23477721]

91. Serences JT, Saproo S. Computational advances towards linking BOLD and behavior. Neuropsychologia. 2011; 50:435–446. [PubMed: 21840553]

92. Naselaris T, et al. Encoding and decoding in fMRI. Neuroimage. 2011; 56:400–410. [PubMed: 20691790]

93. Schoenmakers S, et al. Linear reconstruction of perceived images from human brain activity. Neuroimage. 2013; 83:951–61. [PubMed: 23886984]

94. Cowen AS, et al. Neural portraits of perception: reconstructing face images from evoked brain activity. Neuroimage. 2014; 94:12–22. [PubMed: 24650597]

95. Kok P, et al. Prior expectations bias sensory representations in visual cortex. J. Neurosci. 2013; 33:16275–84. [PubMed: 24107959]

96. Scolari M, et al. Optimal Deployment of Attentional Gain during Fine Discriminations. J. Neurosci. 2012; 32:1–11. [PubMed: 22219264]

97. Ester EF, et al. A Neural Measure of Precision in Visual Working Memory. J. Cogn. Neurosci. 2013 DOI: 10.1162/jocn_a_00357.

98. Anderson DE, et al. Attending Multiple Items Decreases the Selectivity of Population Responses in Human Primary Visual Cortex. J. Neurosci. 2013; 33:9273–9282. [PubMed: 23719796]

99. Byers A, Serences JT. Enhanced attentional gain as a mechanism for generalized perceptual learning in human visual cortex. J. Neurophysiol. 2014; 112:1217–27. [PubMed: 24920023]

100. Kok P, de Lange FP. Shape Perception Simultaneously Up- and Downregulates Neural Activity in the Primary Visual Cortex. Curr. Biol. 2014; 24:1531–1535. [PubMed: 24980501]

101. Sprague TC, et al. Reconstructions of Information in Visual Spatial Working Memory Degrade with Memory Load. Curr. Biol. 2014 DOI: 10.1016/j.cub.2014.07.066.

102. Tong F, Pratte MS. Decoding Patterns of Human Brain Activity. Annu. Rev. Psychol. 2012; 63:483–509. [PubMed: 21943172]

103. Pouget A, et al. Inference and computation with population codes. Annu. Rev. Neurosci. 2003; 26:381–410. [PubMed: 12704222]

104. Ma WJ, et al. Bayesian inference with probabilistic population codes. Nat. Neurosci. 2006; 9:1432–8. [PubMed: 17057707]

105. Pouget A, et al. Information processing with population codes. Nat. Rev. Neurosci. 2000; 1:125–32. [PubMed: 11252775]

106. Naselaris T, et al. Bayesian Reconstruction of Natural Images from Human Brain Activity. Neuron. 2009; 63:902–915. [PubMed: 19778517]

107. Nishimoto S, et al. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Curr. Biol. 2011; 21:1641–1646. [PubMed: 21945275]

108. Cohen MR, Maunsell JHR. Attention improves performance primarily by reducing interneuronal correlations. Nat. Neurosci. 2009; 12:1594–1600. [PubMed: 19915566]

109. Cohen MR, Maunsell JHR. A neuronal population measure of attention predicts behavioral performance on individual trials. J. Neurosci. 2010; 30:15241–53. [PubMed: 21068329]

110. Cohen MR, Maunsell JHR. Using Neuronal Populations to Study the Mechanisms Underlying Spatial and Feature Attention. Neuron. 2011; 70:1192–1204. [PubMed: 21689604]

111. DiCarlo JJ, et al. How does the brain solve visual object recognition? Neuron. 2012; 73:415–34. [PubMed: 22325196]

112. Pagan M, Rust NC. Dynamic target match signals in perirhinal cortex can be explained by instantaneous computations that act on dynamic input from inferotemporal cortex. J. Neurosci. 2014; 34:11067–84. [PubMed: 25122904]

113. Saalmann YB, et al. The Pulvinar Regulates Information Transmission Between Cortical Areas Based on Attention Demands. Science (80-.). 2012; 337:753–756.

114. Gregoriou GG, et al. Long-range neural coupling through synchronization with attention. Prog. Brain Res. 2009; 176:35–45. [PubMed: 19733748]

115. Miller EK, Buschman TJ. Cortical circuits for the control of attention. Curr. Opin. Neurobiol. 2013; 23:216–22. [PubMed: 23265963]

116. Peters AJ, et al. Emergence of reproducible spatiotemporal activity during motor learning. Nature. 2014; 510:263–7. [PubMed: 24805237]

117. Sohya K, et al. GABAergic neurons are less selective to stimulus orientation than excitatory neurons in layer II/III of visual cortex, as revealed by in vivo functional Ca2+ imaging in transgenic mice. J. Neurosci. 2007; 27:2145–9. [PubMed: 17314309]

118. Masamizu Y, et al. Two distinct layer-specific dynamics of cortical ensembles during learning of a motor task. Nat. Neurosci. 2014; 17:987–94. [PubMed: 24880217]

119. Agarwal G, et al. Spatially distributed local fields in the hippocampus encode rat position. Science (80-.). 2014; 344:626–30.

120. Khodagholy D, et al. NeuroGrid: recording action potentials from the surface of the brain. Nat. Neurosci. 2014; 18:310–315. [PubMed: 25531570]

121. Huth AG, et al. A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. Neuron. 2012; 76:1210–1224. [PubMed: 23259955]

122. Zénon A, Krauzlis RJ. Attention deficits without cortical neuronal deficits. Nature. 2012

123. Krauzlis RJ, et al. Attention as an effect not a cause. Trends Cogn. Sci. 2014; 18:457–464. [PubMed: 24953964]

124. Ho T, et al. The Optimality of Sensory Processing during the Speed–Accuracy Tradeoff. J. Neurosci. 2012; 32:7992–8003. [PubMed: 22674274]

125. Haak KV, et al. Connective field modeling. Neuroimage. 2012; 66:376–384. [PubMed: 23110879]

126. Seriès P, et al. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. Nat. Neurosci. 2004; 7:1129–35. [PubMed: 15452579]

127. Zhang K, Sejnowski TJ. Neuronal tuning: To sharpen or broaden? Neural Comput. 1999; 11:75–84. [PubMed: 9950722]

128. Pouget A, et al. Narrow versus wide tuning curves: What's best for a population code? Neural Comput. 1999; 11:85–90. [PubMed: 9950723]

129. Ruff DA, Cohen MR. Attention can either increase or decrease spike count correlations in visual cortex. Nat. Neurosci. 2014; 17:1591–7. [PubMed: 25306550]

130. Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. Nat. Neurosci. 2011; 14:811–9. [PubMed: 21709677]

131. Mitchell JF, et al. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron. 2009; 63:879–88. [PubMed: 19778515]

132. Moreno-Bote R, et al. Information-limiting correlations. Nat. Neurosci. 2014; 17:1410–1417. [PubMed: 25195105]

133. Goris RLT, et al. Partitioning neuronal variability. Nat. Neurosci. 2014; 17:858–65. [PubMed: 24777419]

134. Silver MA, Kastner S. Topographic maps in human frontal and parietal cortex. Trends Cogn. Sci. 2009; 13:488–495. [PubMed: 19758835]

135. Wandell B, et al. Visual field maps in human cortex. Neuron. 2007; 56:366–383. [PubMed: 17964252]

136. Arcaro MJ, et al. Retinotopic organization of human ventral visual cortex. J. Neurosci. 2009; 29:10638–52. [PubMed: 19710316]

Highlights

Brains are capacity-limited due to a finite number of neurons and stochastic synapses Attention enhances relevant information at the expense of irrelevant information Multiple attentional modulations are seen at the single neuron or voxel level Population-based stimulus reconstructions index net attention effects on neural codes

More processing, less overall information, information preserved about attended features
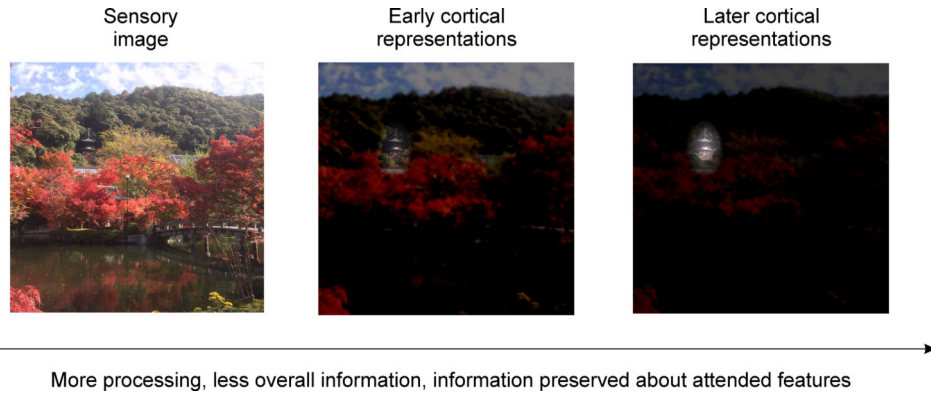
Figure 1. Attention filters behaviorally-relevant information

When viewing a complex natural scene (left), visual processing by a noisy neural system will necessarily result in an overall information loss. If your eyes were fixated on the center of the image, but you were directing attention to the temple nestled among the trees near the top, information about the attended temple would be selectively preserved from degradation by noisy neural processing such that, even at successively later stages of computation, information about the attended location and/or features of the image is still maintained, despite substantial loss of information about unattended components of the image (right panel).
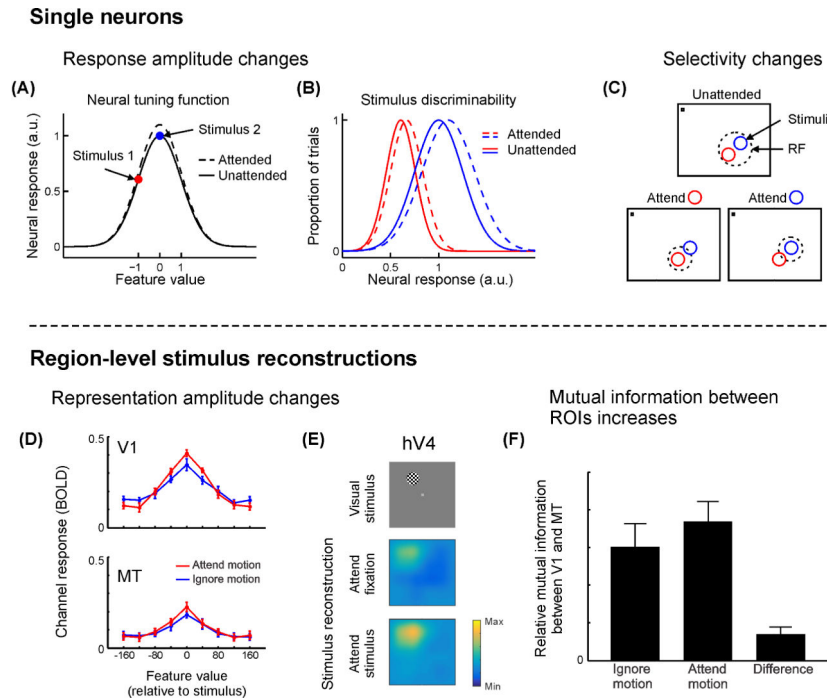
Single neurons

Response amplitude changes

(A) Neural tuning function

(B) Stimulus discriminability

Selectivity changes

(C)

Region-level stimulus reconstructions

Representation amplitude changes

(D)

Mutual information between ROIs increases

(E) hV4

(F)

Figure 2. Attention improves the information content of small- and large-scale neural codes
When attention is directed towards a stimulus presented to a feature-selective neuron, several types of responses are commonly observed. (**A**) Response amplitudes often increase, which increases the dynamic range of the response, and accordingly improves the ability of the neuron to discriminate between two stimuli (**B**). This increased dynamic range enables improved discrimination of multiple stimulus feature values. (**C**) Many neurons show changes in RF properties with attention such that the spatial profile of their response is focused around an attended stimulus placed inside the RF. In contrast, population-level stimulus reconstructions (Box 2) enable assessment of the net impact of all unit-level response changes with attention. (**D**) When participants are instructed to attend to the direction of motion of a moving dot stimulus (as opposed to its contrast), the amplitude of motion direction-selective responses increases in both V1 and MT. (**E**) When participants attend to a flickering checkerboard disc, the reconstructed stimulus image (Box 2) is higher-amplitude compared to when they attend to the fixation point, especially in extrastriate visual ROIs such as human area V4 (hV4). (**F**) When motion is attended (see **D**), the mutual information between V1 and MT is greater compared to when stimulus contrast is attended, suggesting that attention maximizes the transfer of relevant information between brain regions at a population level. Panels A-C cartoon examples. Panels D & F adapted from [9] with permission from the Society for Neuroscience, panel E adapted from [76] with permission from Nature Publishing Group, colormap adjusted.
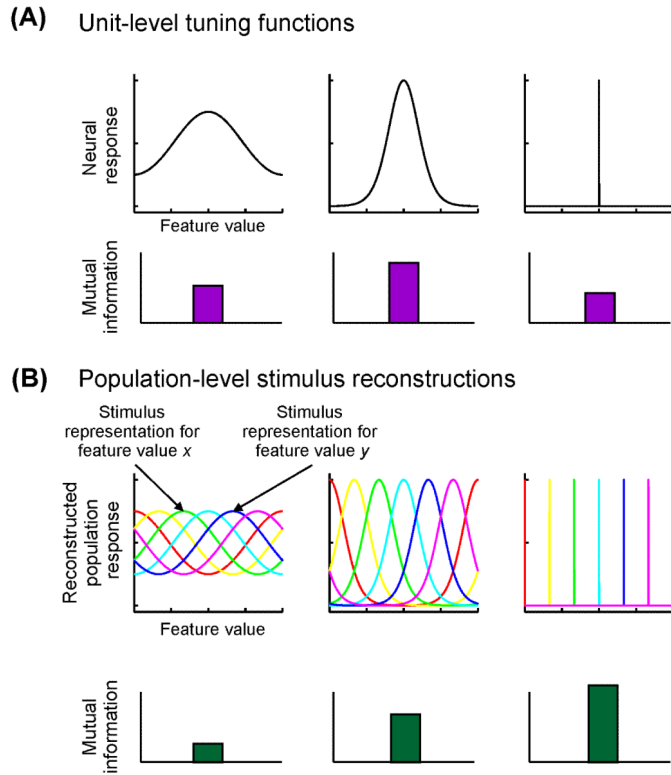
(A) Unit-level tuning functions

(B) Population-level stimulus reconstructions

Figure 3. Information content of units and populations
(A) As described in Box 1, the mutual information (MI) between a unit's response strength and the associated stimulus value is a non-monotonic function of its tuning bandwidth. A non-selective unit will have very low MI (as a result of little variability in response associated with variability in the stimulus feature), but so will a highly selective unit, because it has lower overall entropy. The particular selectivity bandwidth for which a unit has greatest MI about a stimulus feature depends on the shape of the tuning function (TF), noise properties, and the relative frequency of occurrence of different feature values. **(B)** In contrast, for population-level stimulus reconstructions, a narrower reconstructed stimulus representation is more informative, as it reflects a greater level of discriminability between different stimulus feature values. Plotted are cartoon reconstructions of different values of a stimulus. Each color corresponds to a different feature value, and each point along each curve corresponds to the reconstructed activation of the corresponding population response (as in Figs. 2 D-E).
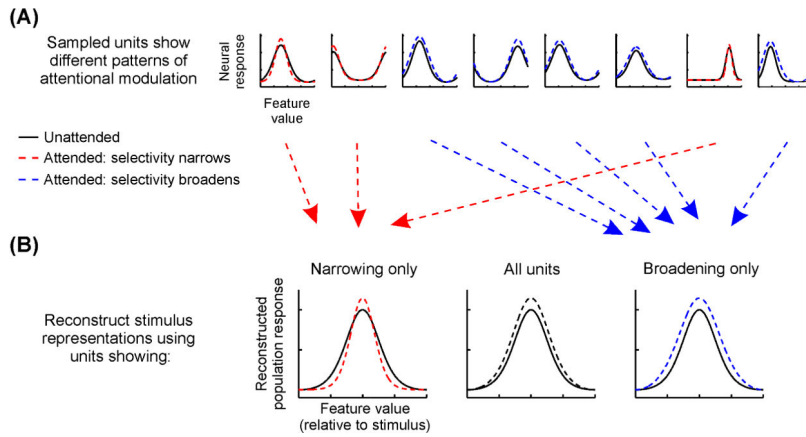
Figure 4. Using stimulus reconstructions to exploit and understand the net impact of heterogeneous response modulations

(A) Attentional modulations of different units are often heterogeneous, reflecting combinations of amplitude increases, baseline changes, and selectivity changes. Thus, even though the mean attentional modulations across units often point in the same direction across studies, there is substantial variability within a given sample of neurons (as shown here in simulated cartoon neural TFs), and any information that is encoded by this variability is usually ignored. **(B)** When using the inverted encoding model (IEM) technique, which combines across modulations across all constituent units, it is possible to ascertain how different types of unit-level response modulations may contribute to stimulus reconstructions by selecting measurement units post hoc that exhibit one type or another of response modulation (e.g., only bandwidth narrowing or broadening) to compute stimulus reconstructions.
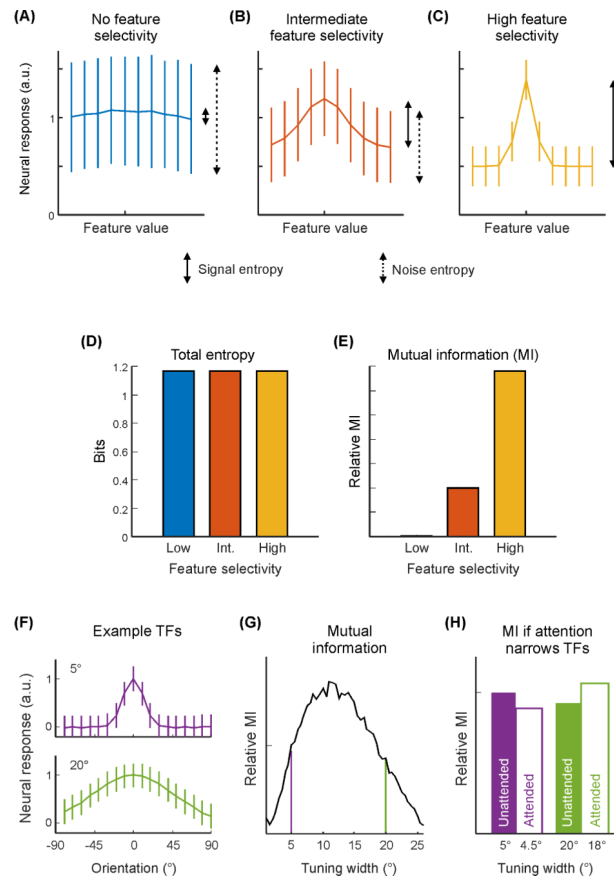
Figure I. Comparison of unit-level information content (mutual information)

For a cartoon neural tuning function (TF; measured in arbitrary units, a. u.) with no **(A)**, intermediate **(B)** or high **(C)** feature selectivity, the overall entropy of the neural response (given equal stimulus probability) is the same **(D)** but the mutual information (MI) between the neural response and the feature value **(E)** is much higher when the neuron exhibits selectivity (e.g., C) than when it is non-selective (e.g., A), with intermediate selectivity (e.g., B) falling in between. In cases where attention narrows the selectivity of a unit, whether such a change results in improved information content depends on the original selectivity before narrowing [127,128]. **(F)** Cartoon of neural TFs for stimulus orientation with different tuning widths (top: narrow tuning width/high selectivity; bottom: large tuning width/low selectivity). **(G)** When MI is plotted as a function of tuning width, any changes in tuning width that are associated with attention can either increase or decrease MI depending on the original tuning width. Vertical lines indicate tuning width values of example TFs in (F). **(H)** For a modest attention-related narrowing in tuning (e.g. ~10% reduction in tuning width), a unit with a small tuning width may actually carry less information with attention, while a unit with a large tuning width would carry more information. Note: panels **F-H** do not equate total entropy across tuning widths (as in panels **A-C**), accounting for differences in MI plots between panel **E** and panels **G-H**.

Encoding models: characterize single measurement unit selectivity
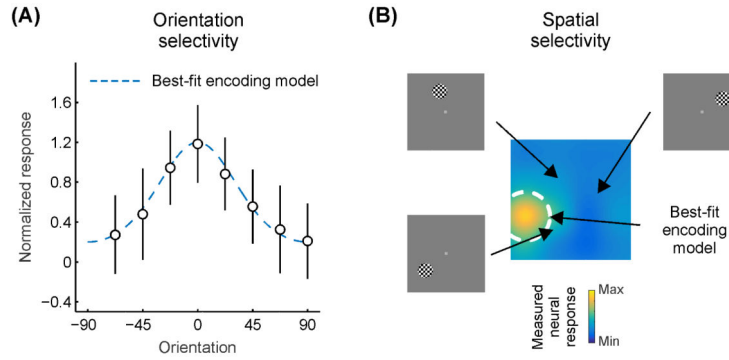
Reconstruction: combine encoding models to reconstruct stimulus and characterize information content
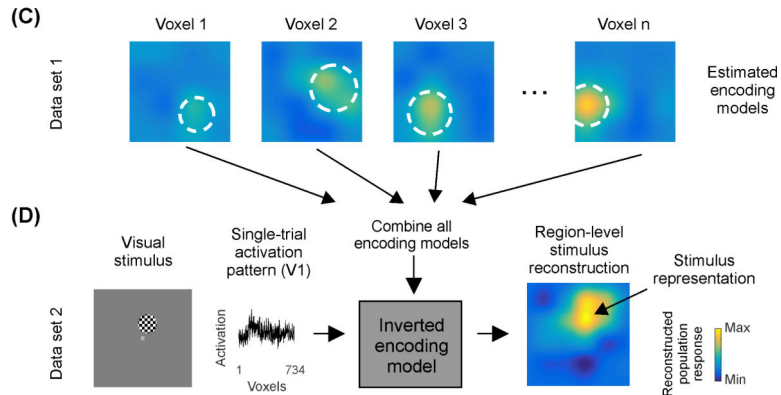
Figure II. Encoding models and stimulus reconstruction

Measurements of neural responses to different stimulus feature values (such as orientation, **(A)**, or spatial position, **(B)**) often reveal selectivity for particular feature values (such as orientation TFs or spatial RFs). Such selectivity can be observed with single-unit firing rates, calcium transients, fMRI BOLD responses, or even scalp EEG. When encoding models are measured for many neurons/voxels/electrodes, it is possible to combine all encoding models to compute an inverted encoding model (IEM, **C**). This IEM allows a new pattern of activation measured using a separate dataset to be transformed into a stimulus reconstruction **(D)**, reflecting the population-level representation of a stimulus along the features spanned by the encoding model (here, visual spatial position). This reconstruction (right) reflects data from a single trial, which is inherently noisy. However, when many similar trials are combined (Fig. 2E), high-fidelity stimulus representations can be recovered.