Properties of V1 neurons tuned to conjunctions of visual features: application of the V1 saliency hypothesis to visual search behavior

Li Zhaoping[1,2], Li Zhe[2]

Abstract

From a computational theory of V1, we formulate an optimization problem to investigate neural properties in the primary visual cortex (V1) from human reaction times (RTs) in visual search. The theory is the V1 saliency hypothesis that the bottom-up saliency of any visual location is represented by the highest V1 response to it relative to the background responses. The neural properties probed are those associated with the less known V1 neurons tuned simultaneously or conjunctively in two feature dimensions. The visual search is to find a target bar unique in color (C), orientation (O), motion direction (M), or redundantly in combinations of these features (e.g., CO, MO, or CM) among uniform background bars. A feature singleton target is salient because its evoked V1 response largely escapes the iso-feature suppression on responses to the background bars. The responses of the conjunctively tuned cells are manifested in the shortening of the RT for a redundant feature target (e.g., a CO target) from that predicted by a race between the RTs for the two corresponding single feature targets (e.g., C and O targets). Our investigation enables the following testable predictions. Contextual suppression on the response of a CO-tuned or MO-tuned conjunctive cell is weaker when the contextual inputs differ from the direct inputs in both feature dimensions, rather than just one. Additionally, CO-tuned cells and MO-tuned cells are often more active than the single feature tuned cells in response to the redundant feature targets, and this occurs more frequently for the MO-tuned cells such that the MO-tuned cells are no less likely than either the M-tuned or O-tuned neurons to be the most responsive neuron to dictate saliency for an MO target.

Introduction

Background on visual attention, saliency, and their neural substrates

Spatial visual selection, often called spatial attentional selection, enables vision to select a visual location for detailed processing using limited cognitive resources[1]. It can be generated by goal-dependent (or top-down) mechanisms, such as when we direct our gaze to a book while reading, or by goal-independent (or bottom-up) mechanisms such as when we are distracted from reading by a sudden appearance of something in visual periphery. In this paper, an input is said to be salient when it strongly attracts attention by bottom-up mechanisms, and the degree of this attraction is defined as saliency. Saliency of a visual location is often measured by the speed of a visual search to find a target at this location[2], or by its attentional (exogenous) cueing effect (i.e., the degree it speeds up and/or improves visual discrimination of a probe presented immediately after the brief appearance of the salient cue)[3; 4].

It has been proposed that the primary visual cortex (V1) is responsible for computing saliency[5; 6]. Although this V1 saliency hypothesis is a significant departure from traditional psychological theories[2; 7; 8; 1], in which the neural substrates are not their main concern, it has received substantial support[9; 10; 11; 12; 13; 14; 15]. In particular, behavioral data confirmed an unexpected prediction that an eye of origin singleton (e.g., an item uniquely shown to the left eye among other items shown to the right eye) that is hardly distinctive from other visual inputs can attract attention and gaze qualitatively just like a salient and highly distinctive orientation singleton does — in fact observations[13; 15] show that the eye of origin singleton can be more salient than an orientation singleton. This finding provides a hallmark of the saliency map in V1, because the eye of origin feature is not explicitly represented in any visual cortical area except V1. Functional magnetic resonance imaging and event related potential measurements also confirmed that, when top-down confounds are avoided, a salient location evokes brain activations in V1 but not in the parietal and frontal regions[14], which are thought to be involved in saliency by traditional views[1].

In another study, Koene and Zhaoping[10] measured RTs for finding a target bar unique in color (C), orientation (O), motion direction (M), or redundantly in two of these features (CO, MO, or CM) among background bars which are identical to each other in all features (see Fig. 1). If the RT for a redundant target (e.g., a CO target) is the shorter one of the two RTs for the corresponding single feature targets (e.g., the C and O targets), the RT for the redundant target is said to be the outcome of a race model between the two other RTs (as if to take the RT of the winner in a race between two racers)[16; 17; 18; 19]. Since RTs are stochastic, the RT from a race model is also said to be the result of a statistical facilitation between the RTs of the individual racers. If the RT for a redundant target is shorter than predicted from the statistical facilitation, there is a redundancy gain[20]. According to the V1 saliency hypothesis, the presence or absence of the redundancy gain in behavior should reflect the presence or absence, respectively, of V1 neurons tuned simultaneously to the two visual features distinguishing the redundant target. We call such cells conjunctively tuned CO, MO, or CM cells, each denoted by the feature dimensions in which they are tuned.
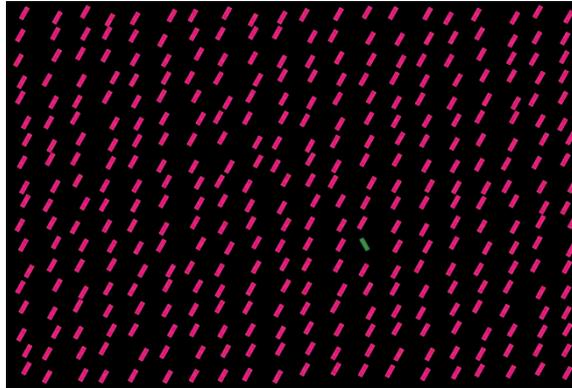
2

Figure 1: A schematic example of the search stimulus by Koene and Zhaoping(10). Data in their behavioral study are used for the current study. Observers searched for a bar unique in color (C), orientation (O), or motion direction (M), or a combination of these features. In this illustration, the target is a double feature CO target, unique in both color and orientation. See Method or the original paper(10) for the actual stimulus details.

Koene and Zhaoping(10) found this redundancy gain for the CO and MO targets but not the CM target, supporting the V1 saliency hypothesis since V1 has CO and MO cells(21; 22; 23), but no CM cells(24).

The finding by Koene and Zhaoping(10) also implies that the extrastriate cortices are unnecessary for the bottom-up saliency of their singleton targets. This is because extrastriate cortices do have the CM conjunctive cells(25; 26), which would have led to a redundancy gain in the CM targets. The implication is consistent with another behavioral observation involving depth cues, which are believed to be processed in extrastriate but not V1(27; 28; 29; 30; 31; 32). It was found that depth cue did not speed up attentional guidance to a target location unless this location was not salient enough to be reported by observers within an RT of one second(33), which is about twice as long as typical RTs to report a feature singleton in Koene and Zhaoping's study. Longer RT events are likely to involve top-down and object/surface recognition processes beyond the bottom-up saliency process (which dominates only in short RT events(34)), and involve extensive neural connections between V1 and extrastriate cortices(35; 36; 37). In addition, the findings by Koene and Zhaoping(10) and others(11) are consistent with the feature combination rule to compute saliency according to the V1 saliency hypothesis. According to this rule, saliency at a location is determined by the highest V1 neural response to that location, without combining responses from multiple neurons responding simultaneously to different input features at the same location. In contrast, the feature combination rules by the traditional saliency models (reviewed by Itti and Koch(1)) compute the saliency value at a location by summing responses to this location from various basic feature maps. Apparently, V1 does not perform any summation across feature dimensions. Hence, higher cortical areas have to be involved if feature summation is to occur for computing a saliency map.
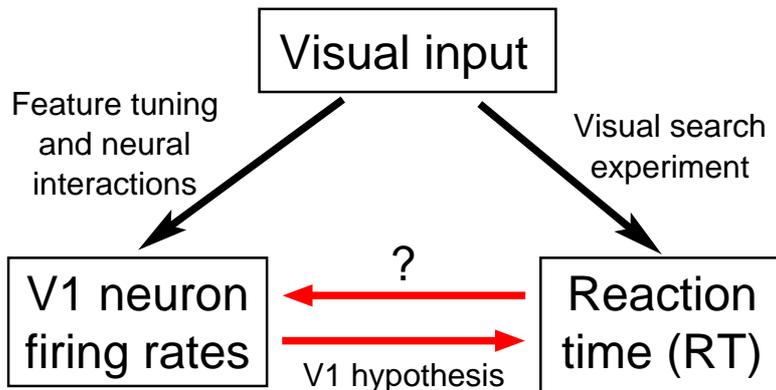
Figure 2: The schematic of our method to probe V1 properties through behavior. Visual inputs drive V1 responses. Meanwhile, the V1 responses determine the behavioral RTs in visual search tasks, according to the hypothesis that the V1 responses represent saliencies of input locations. Therefore, one may probe V1 properties through the relationship between the RTs and V1 responses. In particular, a shorter RT arises from a higher V1 response to the search target relative to those to the background items. Therefore, from the RT data, one can infer relative response levels of the V1 neurons, thereby probing the feature tuning of the V1 neurons and interactions between the neurons.

The goal and the plan for the current study

Whereas the previous studies used known facts about V1 physiology to test, and confirm, the V1 saliency hypothesis, the current study aims to probe the unknown or less known V1 properties assuming that the V1 saliency hypothesis holds (Fig. 2). In particular, Koene and Zhaoping(10) confirmed that the V1 saliency hypothesis is supported by the known facts that V1 contains CO and MO cells but no CM cells. Meanwhile, many physiological properties associated with these conjunctive neurons are less known, or have not been systematically studied. In particular, one would like to ask the following questions. How responsive these conjunctive neurons are compared to the other neurons? How do the intra-cortical interactions between these neurons vary with the feature preferences of the interacting neurons? The current study uses the V1 saliency hypothesis to investigate these less known properties from the behavioral RT data collected by Koene and Zhaoping(10). To do so, we formulate a computational approach based on the V1 saliency hypothesis to solve for aspects of the V1 neural properties from the behavioral RT.

For this study, the theoretical basis is the V1 saliency hypothesis. The hypothesis states that the saliency of a visual location is represented by the highest V1 response to this location relative to the background responses(5; 6). This is regardless of whether this response is from a neuron tuned to orientation (O), color (C), motion (M) direction(38; 39), or other features, or conjunctively to two feature dimensions (e.g. CO or MO)(38; 39; 21; 22; 23). In particular, according to this hypothesis, the most salient location in a scene is the receptive field (RF) of the most activated V1 neuron responding to this scene, regardless of the preferred feature(s) of this neuron. A feature singleton,

such as the search target in Koene and Zhaoping(10) (see Fig. 1), can evoke the highest response to the scene because of a neural property called iso-feature suppression(40). Iso-feature suppression means that V1 neurons tuned to the same or similar features tend to suppress each other's responses via intra-cortical neural connections when their RFs are close to each other(41; 42; 43). For example, a unique vertical bar is very salient in a background of horizontal bars, since different neurons (preferring horizontal orientation) responding to different and neighboring horizontal bars suppress each other by iso-orientation suppression(42), while the neuron (preferring vertical orientation) responding to the unique vertical bar escapes such suppression. Iso-color(44) and iso-motion-direction(45) suppressions are other known examples of iso-feature suppression. To make the highest response to the feature singleton target sufficiently higher than those to the background bars, the following two conditions are required. First, the intra-cortical interactions are sufficiently feature specific such that iso-feature suppression is only substantial between two neurons whose preferred feature(s) are sufficiently similar. (In principle, it should also work if the iso-feature suppression is much stronger when the two neurons prefer sufficiently similar feature(s) than otherwise.) Second, the input feature preference of the neurons should sufficiently differentiate the target and background features. We call these two elements *feature tuning of intra-cortical interactions* and *feature tuning of individual neurons* respectively. Sometimes, *feature tuning of individual neurons* is also referred to as *feature tuning of input preferences.*

Usually the feature singleton search target in Koene and Zhaoping(10) evokes responses from many cells tuned to different features. Some of these cells are tuned to color (C), orientation (O), or motion direction (M), and some are tuned to conjunctions (e.g., CO, MO) of them. We call a neuron a C, O, or M neuron if it is tuned in a single corresponding feature dimension, and a CO, MO, or CM neuron if it is tuned conjunctively in the two corresponding feature dimensions. According to the V1 saliency hypothesis, the highest response among the responses (to the target) from all neurons determines the saliency of the target. This saliency in turn determines the RT to find the target. For example, for a color singleton, a C cell's response is expected or assumed (see Discussion) as most likely to dictate its saliency. Meanwhile, for a CO singleton target in Fig. 1, the dictating response could come from a C, O, or a CO cell, depending on the feature tunings of these cells and of the intra-cortical interactions. We will show that some aspects of these neural properties can be revealed from the RT data through the solution of an optimization problem formulated from the V1 saliency hypothesis.

Previous works(5; 40; 46; 6; 47) have introduced a V1 model to simulate and analyze the intra-cortical mechanisms in order to understand the neural mechanisms behind the V1 saliency hypothesis. We like to point out that the current work probes the V1 neural properties using the V1 saliency hypothesis, the theory, rather than this V1 model. The theory presents a hypothesis about the functional role of the V1 responses, and states that the intra-cortical mechanisms serve to highlight V1 responses to conspicuous locations where input statistics deviates from translation invariance(5; 40; 6). In contrast, by simulating the mechanisms in V1 that give rise to these responses, the model tests whether it is feasible that V1 responses might play the hypothesized role.

5

For simplicity, this V1 model, or model V1, has so far included only model neurons tuned to orientation, except in two examples in which model neurons tuned to color or color-orientation conjunctions are also included(6; 9). However, the theoretical hypothesis is general regarding input feature dimensions and neural mechanisms as it refers to the real, physiological, V1, rather than the simplistic and inaccurate model V1. Indeed, various behavioral tests of the hypothesis have included both the modeled and not modeled feature dimensions: orientation, color, motion direction, and ocular origin(9; 10; 11; 12; 13; 14), since the model V1 is unnecessary when the physiological V1 in human observers are available for these behavioral experiments. Similarly, our formulation, method, and results in the current study depend only on the V1 saliency hypothesis and the general knowledge about the physiological V1, and not on the model.

Our predicted V1 properties from applying the V1 saliency hypothesis to the behavioral data can serve two purposes. First, they can motivate physiological experiments to test the predictions, thus providing further test of the V1 saliency hypothesis. Second, they enable the use of a computational theory as a tool to investigate physiological properties from behavioral data without physiological experiments. We will discuss the implication of our findings in the Discussion.

Method

Behavioral data

The RT data are collected by Koene and Zhaoping(10), which contained all experimental details. In that study, verbal consents from all participants were obtained, as documented by the subject information in the data. The study and the consent procedure were approved by the ethics committee in University College London. Briefly, the search display contained an array of 30×22 colored, tilted, and moving bars. Observers were instructed to find the target bar as soon as possible, and their RTs to find it were measured. There were only two possible iso-luminant colors (green or purple of the same saturation), orientations (left or right tilted from vertical by the same angle), and motion directions (moving to the left or right at the same speed) for all stimulus bars in any search trial. All non-target bars were identical to each other in color, orientation, and motion direction (see Fig. 1), and the target differed from the non-target bars in color, orientation, motion direction, or redundantly in more than one feature dimension. In each search trial, the choices of the target and non-target features were random, and the choice of feature dimension(s) in which the target differed from the non-target was also random. Hence, the possible target conditions included C, O, M, CO, MO, and CM, each defined by the feature dimensions in which the target feature was unique. Each bar was about 1 degree long and 0.2 degree wide. The positions of the bars were randomly jittered from their regular grid locations, such that the horizontal distance between neighboring bars ranged between 1.2 to 3.3 degree and the vertical distance between them ranged between 1.1 to 2.0 degrees. The data considered in this study are from the search trials in which the target bar was at a random location roughly 12.8 degrees from the display center, and at least 11 degrees horizontally. The observers were instructed to press a left or right button as soon as possible for a

target (present in each trial) in the left or right half of the stimulus array, respectively. For a given target condition and a given subject, the mean and standard deviation of the RTs for the correctly performed trials were obtained, and RT outliers are defined as those shorter than 0.2 second or longer than 3 standard deviations from the mean. RTs included for this study exclude the RT outliers and those in trials with an incorrect button press. When the target was unique in color, orientation, or motion direction only, it is called a single feature target; when a target was unique in two features, it is called a double or redundant feature target. For each subject, the orientation and color difference between the target and non-targets, and the motion speed, were roughly pre-adjusted, such that the subject had a mean RT of about 600 ms for each single feature target type. Typically, the average RTs for the double feature targets were around 500 ms as a consequence. Each subject did about 320 search trials for each target type. The percentage of trials excluded in our data analysis, due to button press errors or to the RT being an outlier, is no more than 9.2% (about 5% in average) for each subject in each condition. There were eight subjects, including the authors, Koene and Zhaoping, and six naive subjects.

Extracting properties of the conjunctive cells from behavioral data

For simplicity, we will often narrate as if there is only one cell of each cell type responding to each visual location (or bar) in a search stimulus. This one cell should be understood as the most activated cell of the given cell type. This is because, as far as saliency is concerned, the less activated cells by inputs at a given visual location are irrelevant according to the V1 saliency hypothesis. For the same reason, we often omit an entire cell type when considering neurons and their responses to a visual location, as long as the omitted cells are not the most responsive. For example, to a color singleton target among non-target bars, which have the same orientation as the target bar, the dominating responses are most likely from the C cells rather than the O cells, which are suppressed by iso-orientation suppression. In such a case, the analysis will often omit mentioning the O cells at all.

Linking V1 responses with search RT

Due to iso-feature suppression, the most activated V1 neuron to the search stimuli is most likely the ones responding to the target. For example, a C neuron preferring green will respond most vigorously to a green singleton target among purple distractors. Meanwhile, the population responses to non-targets should be approximately those evoked by a stimulus identical to the search stimulus except for replacing the target by a non-target bar. The level of this population response pattern should be independent of whether this uniform group of bars are green or purple (of the same luminance and saturation), left or right tilted (by the same angle from vertical), and moving to the left or right (at the same speed). Therefore, we make the approximation to view the level of the population responses to the non-targets as independent of the target conditions. Consequently, within the class of the search stimuli in our analysis, the target's saliency is a monotonic function of the highest
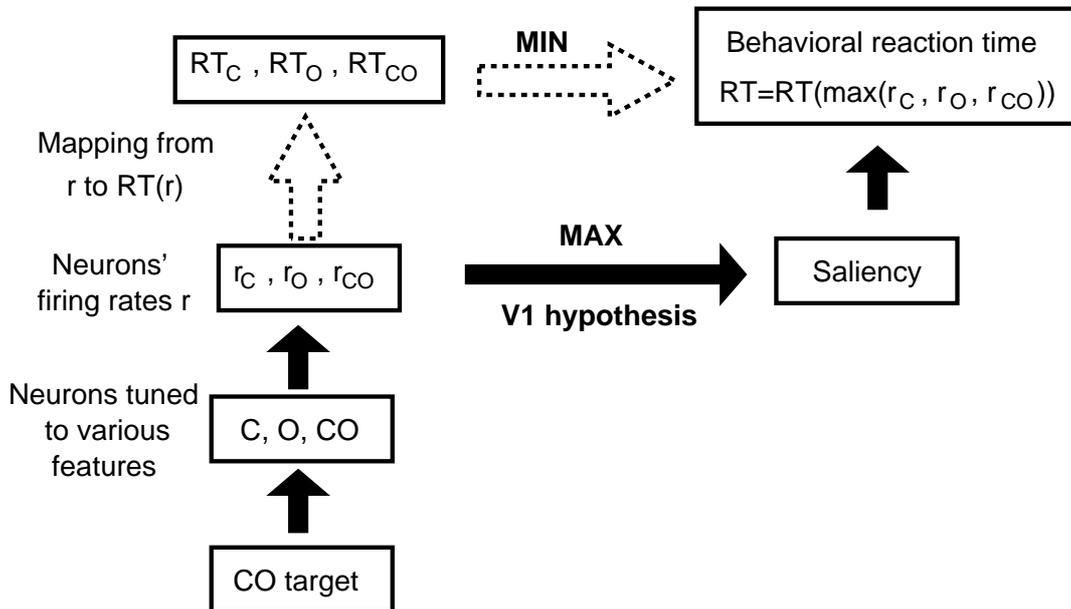
Figure 3: A schematic of the relationship between V1 responses and search RTs. In this example, a CO target activates three types of V1 neurons, tuned to C, O, and CO respectively. Their responses, r_C, r_O, and r_{CO} , are influenced by intra-cortical mechanisms in V1. Their maximum $r_{max} = \max(r_C, r_O, r_{CO})$ determines the target's saliency. Thus the behavioral RT is a function of r_{max} (with distractor responses normalized to 1), through a monotonically decreasing mapping $RT(r_{max})$ determined by the brain mechanisms for saliency read-out and ocular-motor functions. Equivalently, the behavioral RT is $\min(RT_C, RT_O, RT_{CO})$, as the result of a race between the racers C, O, and CO, whose RTs are, respectively, $RT_C = RT(r_C)$, $RT_O = RT(r_O)$, and $RT_{CO} = RT(r_{CO})$.

response r evoked by the target. Since a more salient target leads to a shorter RT by definition, a higher response r maps monotonically to a shorter RT by a mapping $RT(r)$, which depends on the mechanisms of saliency read-out and ocular-motor functions. For example, let r_C be the highest response from the C cells to the color singleton target. Given an observed $RT(r_C)$ to find the target, one can infer the unobserved neural response r_C if the mapping $RT(r)$ is known. Stochastic nature of the neural system gives a distribution of $RT(r_C)$ from many search trials, arising from a corresponding distribution of r_C's.

Obtaining properties of conjunctive cells from the RTs using a race model

Throughout the rest of the Method section, we often use a CO target as an example to derive and illustrate how to probe neural properties, e.g., the relative levels of neural responses r_{CO}, associated with conjunctively tuned cells. The methods and arguments apply analogously to the cases of other double feature targets and neurons.

The saliency for a CO target is not necessarily dictated by r_{CO}, the response from a CO cell, but by the maximum of r_{C}, r_{O}, and r_{CO}, the responses from the C, O, and CO cells, respectively. In other words, the RT for the target is $RT_{\text{target}} = RT(\max(r_{\text{C}}, r_{\text{O}}, r_{\text{CO}}))$, where $\max(...)$ denotes the maximum of the arguments. This method to obtain the RT for a double feature target is called a race model(16; 17; 18; 19; 20), which intuitively assigns as the RT for the target the winning RT in a race between three racers whose respective RTs are $RT(r_{\text{C}})$, $RT(r_{\text{O}})$, and $RT(r_{\text{CO}})$ (Fig. 3, 4). For notational convenience, $RT(r_{\text{C}})$, $RT(r_{\text{O}})$, and $RT(r_{\text{CO}})$ are also denoted as RT_{C}, RT_{O}, and RT_{CO}, respectively. Therefore (Fig. 3),

$$RT_{\text{C}} = RT(r_{\text{C}}) \tag{1}$$

$$RT_{\text{O}} = RT(r_{\text{O}}) \tag{2}$$

$$RT_{\text{CO}} = RT(r_{\text{CO}}) \tag{3}$$

$$RT_{\text{target}} = RT(\max(r_{\text{C}}, r_{\text{O}}, r_{\text{CO}}))$$
$$= \min(RT_{\text{C}}, RT_{\text{O}}, RT_{\text{CO}}), \tag{4}$$

where $\min(...)$ denotes the minimum of the arguments.

The neural activities r_{C} (or r_{O}) are assumed to follow the same probability distribution whether the target is a single feature C (or O) target or a double feature CO target. Hence, the probability distributions of RT_{C}, RT_{O}, and RT_{target} are sampled by behavioral RT data from C, O, and CO target trials respectively. Additionally, r_{C}, r_{O}, and r_{CO} are assumed to be randomly and independently drawn from their respective distributions. Consequently, RT_{C} RT_{O} and RT_{CO} are also randomly and independently drawn from their respective distributions. Meanwhile, RT_{CO}, which cannot be measured behaviorally, can be inferred from other behavioral data. For example, if a RT_{target} sample is shorter than all samples of RT_{C} and RT_{O}, it is likely to represent an underlying RT_{CO} sample according to equation (4). More generally, even when a RT_{target} is not shorter than all samples of RT_{C} and RT_{O}, it is still possible to represent a RT_{CO} sample if its occurrence is more likely than expected from random races between only two racers with RT_{C} and RT_{O} respectively. More formally, an optimization method (see a later section on technical details) can be used to infer the underlying distributions of RT_{C} RT_{O} and RT_{CO} from the behavioral RT samples. Since a monotonic function relates r and RT, relative activity levels among r_{C}, r_{O}, and r_{CO} can then be inferred from the relative values among RT_{C} RT_{O} and RT_{CO}, even though the exact form of the mapping from r to $RT(r)$ is not known and is subject dependent.

Obtain the impacts or contributions of different cells in visual search

The contribution of a neuron to the saliency of a double feature target can be obtained even if the neural activities are not absolutely known. For a CO target, for example, the contribution of the CO cells to the target's saliency is defined as the probability that the CO cell gives the highest evoked
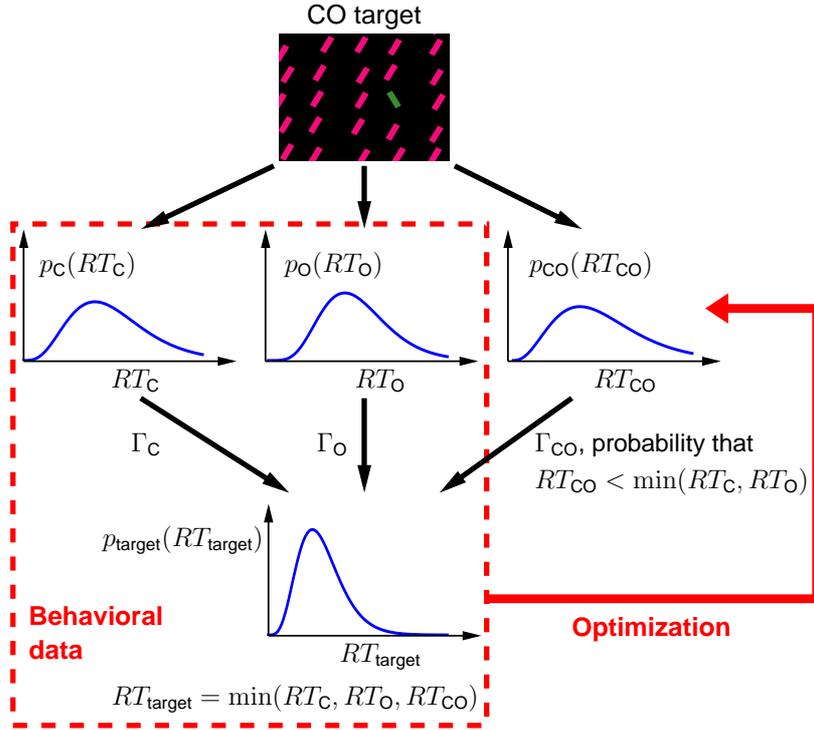
Figure 4: Deriving contributions by various V1 neurons to a double feature target's saliency from the RT. This is illustrated by the example of a CO double feature target. The stochastic V1 responses r_C, r_O, and r_{CO} lead to stochastic $RT_C = RT(r_C)$, $RT_O = RT(r_O)$, and $RT_{CO} = RT(r_{CO})$, with probability distributions $p_C(RT_C)$, $p_O(RT_O)$, and $p_{CO}(RT_{CO})$ respectively. The C, O, or CO cell is the winner of the race with probability Γ_C, Γ_O, and Γ_{CO} respectively, giving target $RT_{target} = \min(RT_C, RT_O, RT_{CO})$. Samples from the probability distributions $p_C(RT_C)$, $p_O(RT_O)$, and $p_{target}(RT_{target})$ are measured as the behavioral RT data for targets C, O, and CO, respectively. From these data, the underlying probability distributions $p_{CO}(RT_C)$, $p_{CO}(RT_O)$, and $p_{CO}(RT_{CO})$ can be inferred by an optimization procedure, and the three contributions Γ_C, Γ_O, and Γ_{CO} can then be calculated.

response (or, equivalently, wins the race among the three racers), i.e.,

$$\Gamma_{\text{CO}} \equiv \text{Probability}(r_{\text{CO}} > r_{\text{C}}, r_{\text{O}})$$
$$= \text{Probability}(RT_{\text{CO}} < RT_{\text{C}}, RT_{\text{O}}). \tag{5}$$

Similarly, the contributions from the C and O cells are, respectively,

$$\Gamma_{\text{C}} \equiv \text{Probability}(RT_{\text{C}} < RT_{\text{O}}, RT_{\text{CO}}), \quad \text{and} \tag{6}$$

$$\Gamma_{\text{O}} \equiv \text{Probability}(RT_{\text{O}} < RT_{\text{C}}, RT_{\text{CO}}). \tag{7}$$

In our data analysis, probability distributions of the RTs are described by probabilities of the RTs in discrete time bins. Due to the finite sizes of these time bins, there is a non-zero probability that more than one racer jointly win a race (by being in the same bin), giving $\Gamma_{\text{C}} + \Gamma_{\text{O}} + \Gamma_{\text{CO}} < 1$. However, this does not change our qualitative conclusions.

One can easily imagine that if the mean RT_{CO} is substantially longer than those of RT_{C} and RT_{O}, the contribution Γ_{CO} by the conjunctive CO cell will be likely close to zero. In contrast, if contribution $\Gamma_{\text{CO}} \approx \Gamma_{\text{C}}, \Gamma_{\text{O}}$, the responses by the conjunctive cells are comparable to those by the single feature tuned neurons.

Assessing the significance of the roles of the conjunctive cells

By definition, the Γ_{CO} will never be negative. Meanwhile, the finite numbers of behavioral samples in our data imply that our sampled probabilities $p_{\text{C}}(RT_{\text{C}})$, $p_{\text{O}}(RT_{\text{O}})$, and $p_{\text{target}}(RT_{\text{target}})$ are noisy versions of the actual probabilities. Consequently, a positive contribution Γ_{CO} is likely obtained even if RT_{target} were sampled from the race between RT_{C} and RT_{O} only. We define Γ_{chance} as the chance level Γ_{CO} value obtained by replacing the RT_{target} data by as many trials (as the number of the CO target trials) of this simulated race winner $\min(RT_{\text{C}}, RT_{\text{O}})$ between the two racers using Monte Carlo method(10). We obtained 1000 evaluations of Γ_{chance}, each from a random set of sampled RTs of the race winner. Γ_{CO} is said to be significant if it is larger than 95% of these Γ_{chance} values, i.e., $p < 0.05$.

The verification of consistency and validity of our method

To verify the consistency of our method, we checked after optimization whether $RT_{\text{race3}} \equiv \min(RT_{\text{C}}, RT_{\text{O}}, RT_{\text{CO}})$, the race winner among the three racers C, O, and CO, has the same distribution as that of our behavioral RT_{target} for the CO target. A large difference between these two distributions indicates a poor performance of our optimization method, and consequently, unreliable results and conclusions from the method. This consistency can be quantified by Δ (defined as $\Delta \equiv D/H$, where D is the K-L divergence between $p_{\text{target}}(RT_{\text{target}})$ and $p_{\text{race3}}(RT_{\text{race3}})$, H is the entropy of $p_{\text{target}}(RT_{\text{target}})$, and both RT distributions are discretized by the same time bins for the calculation), such that a $\Delta \ll 1$ indicates a good agreement between the two distributions. Fig. 5 shows the best and worst consistency cases. For all subjects and double feature target conditions, $\Delta < 0.01$, and typically
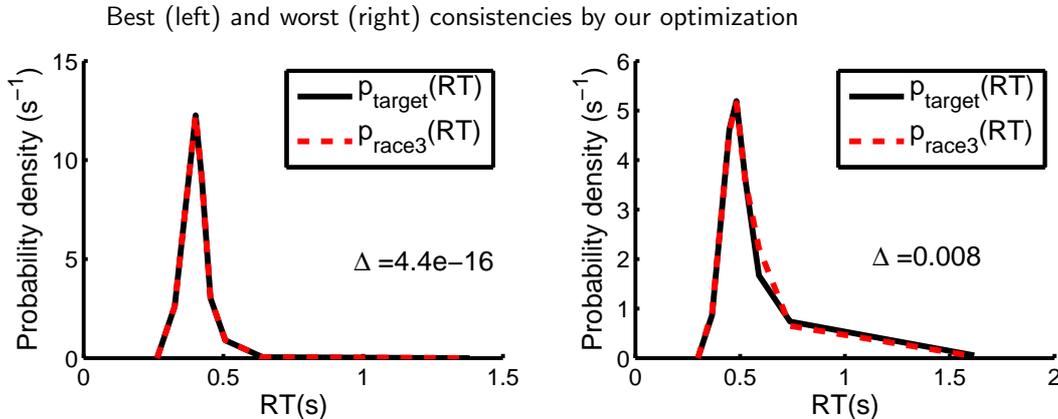
Figure 5: Visualization of the consistency of our method. Shown are the best and the worst consistencies in using our optimization method to probe the double feature tuned cells, among all subjects and all double feature target conditions. A better consistency means a better match between the two curves $p_{\text{target}}(RT_{\text{target}})$ and $p_{\text{race3}}(RT_{\text{race3}})$. Here, $p_{\text{target}}(RT_{\text{target}})$ is the distribution of the behavioral RT data for a double feature target. Meanwhile, $p_{\text{race3}}(RT_{\text{race3}})$ is the corresponding distribution of the winning RT from a race between the three RT racers (e.g., see Fig. 4) whose probability distributions are inferred from the behavioral RT data by our optimization method. In most cases (not shown here), the two curves are not visually distinguishable, similar to that in the plot for the best case.

the curves of $p_{\text{target}}(RT_{\text{target}})$ and $p_{\text{race3}}(RT_{\text{race3}})$ are not visually distinguishable. An analogous Δ value can also be calculated for the probability distributions for any given single feature target, when $p_{\text{target}}(RT_{\text{target}})$ and $p_{\text{race3}}(RT_{\text{race3}})$, respectively, are replaced by the measured and inferred (by the optimization) probability distributions of a given singleton target. For all subjects and all single feature targets, such Δ values are all smaller than 0.005. Hence, our optimization method is highly reliable and gives consistent results.

Meanwhile, our calculated contributions Γ by the various feature tuned cells depend on the number and the placement of the time bins to discretize the RT data. Smaller bins give fewer RT samples in each bin, making the sampled distributions noisier and Γ_{chance} larger. Larger bins give coarser distributions, making it more difficult to distinguish the race winner, since joint winners in a race are more likely. Given the number N of the bins, we place the bins such that each of the first $N - 1$ bins contains roughly the same total number of RT samples from all target types, while the last bin is the reserve for possible long RTs (from the double feature cells) which never wins the race. Given our RT data, when N is between 7-13, such that the probability that the race is won by joint winners is on average between 10-20%, our results do not qualitatively depend on the number N of the bins. This paper shows the results for when $N = 8$.

Technical details in the methods

Optimization method to calculate RT distributions generated by the responses of various types of neurons

For each subject and each target type, the RTs in the N time bins are described by a vector $\mathbf{n} = (n_1, n_2, \ldots, n_N)$, with n_i = the number of RT samples in the ith time bin with $t_{i-1} \leq RT < t_i$. Let $\mathbf{n}_\mathrm{C}$, $\mathbf{n}_\mathrm{M}$, $\mathbf{n}_\mathrm{O}$, $\mathbf{n}_\mathrm{CM\text{-}target}$, $\mathbf{n}_\mathrm{CO\text{-}target}$, and $\mathbf{n}_\mathrm{MO\text{-}target}$ denote these vectors for targets C, M, O, CM-target, CO-target, and MO-target respectively for a given subject. Let the probability distributions $\mathbf{p}_\mathrm{C}$, $\mathbf{p}_\mathrm{M}$, $\mathbf{p}_\mathrm{O}$, $\mathbf{p}_\mathrm{CM}$, $\mathbf{p}_\mathrm{CO}$, and $\mathbf{p}_\mathrm{MO}$ denote the probability of RT_C, RT_M, RT_O, RT_CM, RT_CO and RT_MO respectively in these same time bins. Their likelihood given $\mathbf{n}_\mathrm{C}$, $\mathbf{n}_\mathrm{M}$, $\mathbf{n}_\mathrm{O}$, $\mathbf{n}_\mathrm{CM\text{-}target}$, $\mathbf{n}_\mathrm{CO\text{-}target}$, and $\mathbf{n}_\mathrm{MO\text{-}target}$ is

$$L(\mathbf{p}_\mathrm{C}, \mathbf{p}_\mathrm{M}, \mathbf{p}_\mathrm{O}, \mathbf{p}_\mathrm{CM}, \mathbf{p}_\mathrm{CO}, \mathbf{p}_\mathrm{MO}) \propto \prod_a^N {p_{\mathrm{C}_a}}^{n_{\mathrm{C}_a}} \cdot \prod_b^N {p_{\mathrm{M}_b}}^{n_{\mathrm{M}_b}} \cdot \prod_c^N {p_{\mathrm{O}_c}}^{n_{\mathrm{O}_c}} \cdot$$

$$\prod_d^N {p_{\mathrm{CM\text{-}race3}_d}}^{n_{\mathrm{CM\text{-}target}_d}} \cdot \prod_e^N {p_{\mathrm{CO\text{-}race3}_e}}^{n_{\mathrm{CO\text{-}target}_e}} \cdot \prod_f^N {p_{\mathrm{MO\text{-}race3}_f}}^{n_{\mathrm{MO\text{-}target}_f}} \quad (8)$$

In the above equation, n_{X_i} and p_{X_i} denote the ith element in the vector $\mathbf{n}_\mathrm{X}$ and $\mathbf{p}_\mathrm{X}$ respectively, for X = C, M, O, CM-target, CO-target, MO-target, CM-race3, CO-race3, or MO-race3. Meanwhile, $\mathbf{p}_\mathrm{CM\text{-}race3}$ is the probability distribution of the RTs as the result of a race between three racers whose RTs follow probability distributions $\mathbf{p}_\mathrm{C}$, $\mathbf{p}_\mathrm{M}$, and $\mathbf{p}_\mathrm{CM}$ respectively, i.e.,

$$\begin{aligned}
p_{\mathrm{CM\text{-}race3}_i} = & p_{\mathrm{C}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{M}_j}\Big)\Big(\sum_{k=i+1}^N p_{\mathrm{CM}_k}\Big) + p_{\mathrm{M}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{C}_j}\Big)\Big(\sum_{k=i+1}^N p_{\mathrm{CM}_k}\Big) \\
& + p_{\mathrm{CM}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{C}_j}\Big)\Big(\sum_{k=i+1}^N p_{\mathrm{M}_k}\Big) + p_{\mathrm{C}_i} p_{\mathrm{M}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{CM}_j}\Big) \\
& + p_{\mathrm{C}_i} p_{\mathrm{CM}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{M}_j}\Big) + p_{\mathrm{M}_i} p_{\mathrm{CM}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{C}_j}\Big) + p_{\mathrm{C}_i} p_{\mathrm{M}_i} p_{\mathrm{CM}_i}
\end{aligned} \quad (9)$$

Similarly, the components of $\mathbf{p}_\mathrm{CO\text{-}race3}$ and $\mathbf{p}_\mathrm{MO\text{-}race3}$ are

$$\begin{aligned}
p_{\mathrm{CO\text{-}race3}_i} = & p_{\mathrm{C}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{O}_j}\Big)\Big(\sum_{k=i+1}^N p_{\mathrm{CO}_k}\Big) + p_{\mathrm{O}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{C}_j}\Big)\Big(\sum_{k=i+1}^N p_{\mathrm{CO}_k}\Big) \\
& + p_{\mathrm{CO}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{C}_j}\Big)\Big(\sum_{k=i+1}^N p_{\mathrm{O}_k}\Big) + p_{\mathrm{C}_i} p_{\mathrm{O}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{CO}_j}\Big) \\
& + p_{\mathrm{C}_i} p_{\mathrm{CO}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{O}_j}\Big) + p_{\mathrm{O}_i} p_{\mathrm{CO}_i}\Big(\sum_{j=i+1}^N p_{\mathrm{C}_j}\Big) + p_{\mathrm{C}_i} p_{\mathrm{O}_i} p_{\mathrm{CO}_i}
\end{aligned} \quad (10)$$

$$p_{\text{MO-race3}i} = p_{\text{M}i}\left(\sum_{j=i+1}^{N} p_{\text{O}j}\right)\left(\sum_{k=i+1}^{N} p_{\text{MO}k}\right) + p_{\text{O}i}\left(\sum_{j=i+1}^{N} p_{\text{M}j}\right)\left(\sum_{k=i+1}^{N} p_{\text{MO}k}\right)$$
$$+ p_{\text{MO}i}\left(\sum_{j=i+1}^{N} p_{\text{M}j}\right)\left(\sum_{k=i+1}^{N} p_{\text{O}k}\right) + p_{\text{M}i}p_{\text{O}i}\left(\sum_{j=i+1}^{N} p_{\text{MO}j}\right) \tag{11}$$
$$+ p_{\text{M}i}p_{\text{MO}i}\left(\sum_{j=i+1}^{N} p_{\text{O}j}\right) + p_{\text{O}i}p_{\text{MO}i}\left(\sum_{j=i+1}^{N} p_{\text{M}j}\right) + p_{\text{M}i}p_{\text{O}i}p_{\text{MO}i}$$

We obtain the most likely $\mathbf{p}_{\text{C}}$, $\mathbf{p}_{\text{M}}$, $\mathbf{p}_{\text{O}}$, $\mathbf{p}_{\text{CM}}$, $\mathbf{p}_{\text{CO}}$, and $\mathbf{p}_{\text{MO}}$ by minimizing the negative log likelihood $-\ln L(\mathbf{p}_{\text{C}}, \mathbf{p}_{\text{M}}, \mathbf{p}_{\text{O}}, \mathbf{p}_{\text{CM}}, \mathbf{p}_{\text{CO}}, \mathbf{p}_{\text{MO}})$. For this optimization, we use the "fmincon" function in MATLAB, imposing the constraint that each of $\mathbf{p}_{\text{C}}$, $\mathbf{p}_{\text{M}}$, $\mathbf{p}_{\text{O}}$, $\mathbf{p}_{\text{CM}}$, $\mathbf{p}_{\text{CO}}$, and $\mathbf{p}_{\text{MO}}$ has non-negative components and is normalized, e.g., $\sum_{i=1}^{N} p_{\text{C}i} = 1$.

Quantifying the consistency of our optimization method

To quantify the consistency of our optimization method, we first obtain an unbiased estimation of $\mathbf{p}_{\text{CO-target}}$ of the $RT_{\text{CO-target}}$ as

$$p_{\text{CO-target}i} = \frac{n_{\text{CO-target}i}}{\sum_{j=1}^{N} n_{\text{CO-target}j}}. \tag{12}$$

The difference between $\mathbf{p}_{\text{CO-race3}}$ (as in equation (10) above) and $\mathbf{p}_{\text{CO-target}}$ can be measured by Kullback-Leibler divergence

$$D = \sum_{i=1}^{N} p_{\text{CO-target}i} \ln \frac{p_{\text{CO-target}i}}{p_{\text{CO-race3}i}}. \tag{13}$$

The quality of the consistency of our optimization is quantified by $\Delta = D/H$, where H is the entropy of $\mathbf{p}_{\text{CO-target}}$

$$H = -\sum_{i=1}^{N} p_{\text{CO-target}i} \ln p_{\text{CO-target}i}. \tag{14}$$

Calculating the contributions by various cell types to the saliency of double feature targets

For example, for the CO target, the contributions Γ_{C}, Γ_{O}, and Γ_{CO} are respectively

$$\Gamma_{\text{C}} = \sum_{i=1}^{N} p_{\text{C}i}\left(\sum_{j=i+1}^{N} p_{\text{O}j}\right)\left(\sum_{k=i+1}^{N} p_{\text{CO}k}\right), \tag{15}$$

$$\Gamma_{\text{O}} = \sum_{i=1}^{N} p_{\text{O}i}\left(\sum_{j=i+1}^{N} p_{\text{C}j}\right)\left(\sum_{k=i+1}^{N} p_{\text{CO}k}\right), \tag{16}$$

$$\Gamma_{\text{CO}} = \sum_{i=1}^{N} p_{\text{CO}i}\left(\sum_{j=i+1}^{N} p_{\text{C}j}\right)\left(\sum_{k=i+1}^{N} p_{\text{O}k}\right). \tag{17}$$

The contributions in the case of other double feature targets are obtained analogously.

The policy of placing the time bins

Let $RT_{\min}$ and $RT_{\max}$ be, respectively, the minimum and maximum RTs of a subject regardless of target types. Let $t_{i-1} < t_i$ be the boundaries of the ith time bin containing RTs satisfying $t_{i-1} \leq RT < t_i$. Given the number N of time bins, the boundaries $t_0 < t_1 < ... < t_N$ are chosen such that $t_0 = RT_{\min} - 0.0001$ second, $t_{N-1} = RT_{\max} + 0.0001$ second, $t_N = \infty$, and, if $i < N$, $n_{Ci} + n_{Mi} + n_{Oi} + n_{CM\text{-target}\,i} + n_{CO\text{-target}\,i} + n_{MO\text{-target}\,i} \approx$ constant which does not depend on i. The last time bin bounded by $t_{N-1} < t_N = \infty$ serves as a reservoir for the possibility of the long RT_{CM}, RT_{CO}, and RT_{MO} which never win the races and thus could not be manifested in (or determined by) the behavioral RT data.

Results

For the CO target, Fig. 6A shows the probabilities of RT_C, RT_O, RT_{target} from the behavioral data and that of the inferred RT_{CO} by the optimization for a typical subject. As expected for the RTs of the race winner, the RT_{target} is generally smaller than all the other RTs (of the individual racers). Fig. 6B shows that, for this subject, the saliency of the CO target is determined most likely by the V1 neuron tuned to the O feature, with $\Gamma_O = 0.42$, and least likely by the neuron tuned to the CO feature, with $\Gamma_{CO} = 0.13$. However, Γ_{CO} is significantly larger than Γ_{chance} which is typically around 0.05 for all subjects and double feature target conditions. Hence, for this subject, the CO neuron is so responsive to the CO target that it has a substantial probability of $\Gamma_{CO} = 0.13$ to respond more vigorously than the single feature tuned C and O cells to the CO target.

In Fig. 6A, the distribution of the inferred RT_{CO} is multi-modal, unlike typical RT distributions. This does not mean that our optimization is faulty, as it is caused by the following. First, the race model is better at determining the shorter RT_{CO}s which are more likely to win the race to be manifested as RT_{target}. The longer RT_{CO}s are under-determined and are largely determined by the probability normalization constraint. Meanwhile, these longer RT_{CO}s matter little to Γ_{CO} since they do not win the race. In fact, Fig. 6A omitted the last time bin, which contains no behavioral RT samples for any target types but absorbs the longer RT_{CO}s which never even jointly win the race. Second, our RT data do not allow us to determine how likely it is that the CO cell is the most activated neuron by the C or O target to dictate RT_C or RT_O, respectively. We have thus for simplicity assumed that the CO cells never dictate RT_C or RT_O for the C or O targets respectively. If, however, CO cells do dictate RT_C and RT_O occasionally, the RTs by the C and O neural racers should be longer than those shown in Fig. 6A, and, consequently, some more trials of RT_{target} should be attributed to RT_{CO} to make $p_{CO}(RT_{CO})$ resemble typical RT distributions. The analysis above implies that our inferred Γ_{CO} is in fact the additional contribution by the CO neurons beyond their hidden contribution which has been attributed to the C and O cells for simplicity (see Discussion
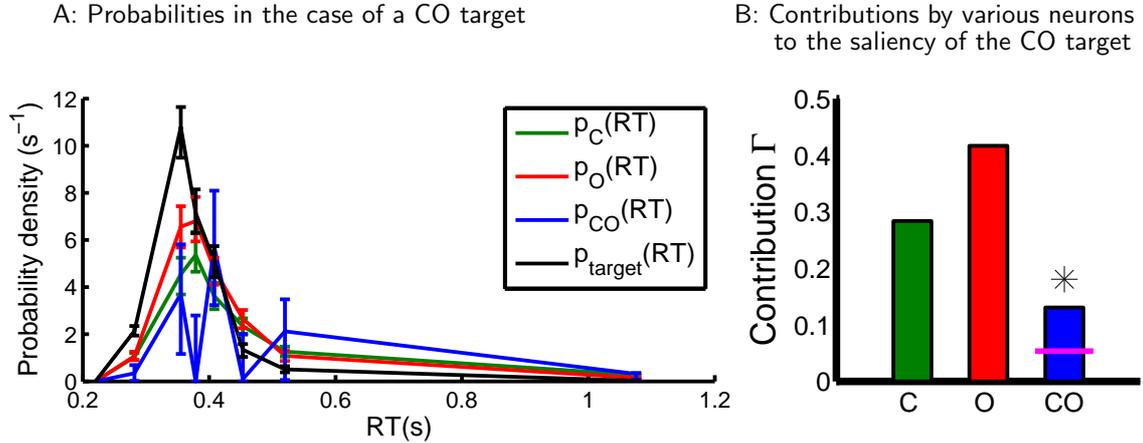
15

A: Probabilities in the case of a CO target

B: Contributions by various neurons to the saliency of the CO target

Figure 6: The results for the CO target from a typical subject. A: probability densities for RT_C, RT_O, RT_{CO}, and RT_{target}. Each density function is plotted as piece-wise lines linking discrete points, with the ith point at $RT = (t_i + t_{i-1})/2$ horizontally and $p_i/(t_i - t_{i-1})$ vertically, where p_i is the probability that the corresponding RT is in the ith time bin ($t_{i-1} \leq RT < t_i$). All curves start at $p_0 = 0$ at t_0. For RT_C, RT_O, and RT_{target}, the probability $p_i = n_i/(\sum_j n_j)$, where n_i is the number of RT samples in the ith time bin for the corresponding target. For RT_{CO}, p_i is from the outcome of the optimization. The error bars are generated as follows. For each target type, let $RT_1 < RT_2 < ...RT_a < ... < RT_M$ be all the behavioral RT samples included, and let the cumulative RT distribution $cdf(RT)$ for this target be approximated by a function which has piece-wise interpolations between discrete functional values $cdf(RT = RT_a) = (a - 0.5)/M$ and has $cdf(RT < RT_1) = 0$ and $cdf(RT > RT_M) = 1$. Randomly generate M simulated RT samples using this $cdf(RT)$. Using these simulated RT samples (as if they were the original RT data) for all target types, we obtain another measurement of the probability densities for all target types and, via our optimization method, all neuron types. Repeat such measurements 500 times. Each error bar has its lower and higher values at the 16th and 84th percentiles, respectively, of the corresponding density measurements. B: Contributions Γ_C, Γ_O, and Γ_{CO} of the C, O, and CO neurons, respectively, to the saliency of the CO double feature target for this subject. Each contribution is the probability that the corresponding neuron dictates the saliency of the CO target (by giving the highest response among responses from all three types of neurons to the CO target). In obtaining their values, probabilities of RT_C and RT_O from optimization outcomes, rather than behavioral data, were used. The '*' on top of Γ_{CO} indicates that Γ_{CO} is significantly different from Γ_{chance} whose mean value is marked by the magenta line.

16

for more details). Third, the probabilities inferred from finite numbers of RT samples are noisy, contributing to the irregularity in the inferred RT_{CO} distribution.

Fig. 7 shows the contributions by various feature tuned neurons to the saliencies of different double feature targets for all subjects. Fig. 7A shows that, among 8 subjects, 5, 7, and 2 subjects have their conjunctive cells contributing significantly to the corresponding double feature targets CO, MO, and CM respectively. A t-test is used to see whether the subject-averaged contribution by any double feature neuron is significantly larger than the subject averaged chance level Γ_{chance}. The answer is affirmative except for the CM cells, confirming the conclusion by Koene and Zhaoping(10) that the behavioral RTs for a double feature CO target or MO target, but not the CM target, is significantly shorter than predicted from a race between the RTs for the two corresponding single feature targets. In addition, the current results reveal quantitatively the impacts of the double feature tuned neurons to the saliencies of the double feature targets, and compare them with the impacts of the single feature tuned neurons. Averaged across subjects, Γ_{MO} is not significantly different from Γ_M and Γ_O, but Γ_{CO} is significantly lower than Γ_O and marginally lower than Γ_C. Hence, the MO cells have a larger impact than the CO cells on the saliency of their corresponding double feature target. In particular, the chance Γ_{MO} for the MO cell to be the highest responding neuron to dictate the saliency of a MO double feature target is no less than that (Γ_M or Γ_O) for either of the single feature tuned M and O cells. Meanwhile, the chance Γ_{CO} for the CO cell to be the highest responding neuron to dictate the saliency of a CO double feature target is substantial, but is only about half of that (Γ_C or Γ_O) for either of the single feature tuned C and O cells. These results will be used to infer the less known properties of the double feature cells in Discussion.

Note that the relative contributions Γ_C, Γ_M, and Γ_O in our results cannot be generally interpreted as relative significance of the roles played by the corresponding single feature tuned cells. For example, if a much smaller orientation contrast between a target and non-targets were employed in our stimuli for O, CO, and MO targets (note that the same orientation contrast was used in these targets by our experimental design), then the saliency of the CO target would be due more to its unique color rather than its unique orientation, and, similarly, the saliency of the MO target would be due more to its unique motion direction rather than its unique orientation. Consequently, the ratio Γ_O/Γ_C for the CO target and the ratio Γ_O/Γ_M for the MO target will be reduced. Nevertheless, our conclusions regarding the contributions by the conjunctive cells relative to those by the single feature tuned cells should not be as sensitive to the exact feature contrasts in the stimuli, since the conjunctive cells have to be more active than both of the corresponding single feature cells to make an impact.

Discussion

Summary of the results and their predictions on V1 physiology

Using RTs in visual search for feature singletons to assess the saliencies of the search targets, and using the V1 saliency hypothesis, this study probes the properties of the less-known V1 cells tuned
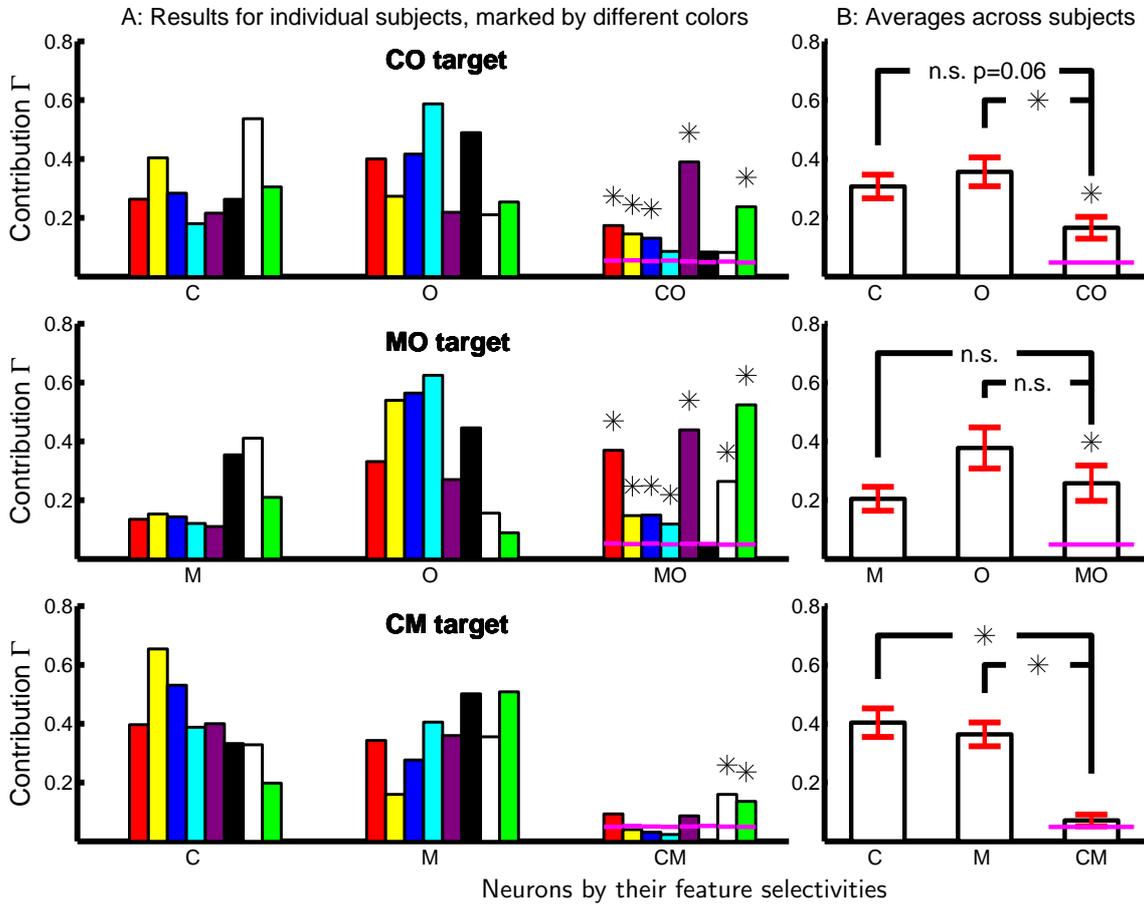
Figure 7: Contributions by various feature-tuned neurons to the three types of double feature targets. Results are shown for individual subjects in A and averaged across subjects in B. The two subjects marked by white and green colored bars are Koene and Zhaoping, experimenters for the behavioral data and the only non-naive subjects. The plots are in the same format as that in Fig. 6B. In B, Γ_{chance} is averaged across subjects, and a subject-averaged contribution by the conjunctive cells is marked as significant if it is significantly different ($p < 0.05$) from this Γ_{chance} by a t-test. The error bars mark the standard errors of the means. An '*' above a bar for the double feature tuned cell (in A or B) indicates that the contribution by this cell (to the saliency of the double feature target) is significantly above the chance level. In B, an '*' or 'n.s.' linking the contribution of a conjunctive cell and that of a single feature tuned cell marks, respectively, a significant or insignificant difference between them (by a matched sample t-test). Qualitatively the same results are obtained in B when data from non-naive subjects, Koene and Zhaoping, are excluded.

conjunctively to more than one feature dimension. We are particularly interested in the activities of the conjunctively tund neurons relative to those of the single feature tuned neurons. These relative activities, when they are sufficiently high, make their impacts on the saliencies of the visual inputs, such that they can shorten the RT to find a double or redundant feature target beyond that predicted by a statistical facilitation between the two corresponding single feature targets. In other words, the relatively higher activities of the conjunctive neurons can be manifested as redundancy gains in the RTs of the double feature targets(20). The relative activities of the conjunctive neurons can be quantified from the redundancy gains by applying the V1 saliency hypothesis. The results show that (1) the chance Γ_{MO} for the MO cell to be the most active neuron in response to a MO double feature target is no less than that (Γ_M or Γ_O) for either of the single feature tuned M and O cells; and (2) the chance Γ_{CO} for the CO cell to be the most active neuron in response to a CO double feature target is substantial but about half of that (Γ_C or Γ_O) for either of the single feature tuned C and O cells. Additionally, our results show that there is no significant chance for the CM cells to be the most active neuron in response to a CM double feature target, suggesting an absence of such neurons in V1, consistent with the previous finding(10) and physiological observations(24).

The impact of the conjunctive cells on the double feature targets predicts that these cells tend to respond to their preferred stimulus more vigorously and experience weaker contextual suppressions when the contextual inputs differ from their preferred stimulus in both, rather than one, feature dimensions. This should be caused by both of the following. One is a sufficient feature tuning of the conjunctive cells in both feature dimensions, and the other is a sufficient feature tuning of the intra-cortical interactions between these cells (or between these cells and the single feature tuned cells). The roles of these two types of feature tunings in saliency are further elaborated next.

Two types of V1 feature tuning properties

V1 saliency hypothesis implies that the highest responses to the feature singletons are higher than those to the uniformly featured non-targets. Mechanistically, this requires the following two components. First, neurons responding to the non-targets should suppress each other by iso-feature suppression, the V1 property that nearby neurons preferring the same or similar feature(s) suppress each other(41; 42; 43). Second, the neuron preferring and responding to the target should largely escape the iso-feature suppression from neurons responding to the non-targets. These two components require two types of feature tunings to be sufficiently strong. One is the feature tuning in the input feature preference of the V1 cells. Cells preferring the target feature should prefer the non-target features much less or not at all. The other is the feature tuning of the intra-cortical interactions(5). It specifies how quickly the intra-cortical suppression decays with the difference between the preferred features of the two interacting neurons. By sufficient feature tuning of the interactions, neurons preferring the non-targets should direct their iso-feature suppression much more to each other than to neurons preferring the target. Sufficient feature tunings in both the input preference of the neurons and interactions between neurons ensures that the neurons most activated by the target should largely escape the iso-feature suppression from the neurons responding to the

non-targets.

Following the analysis above, sufficient feature tuning associated with the conjunctive cells, both in input preference and in intra-cortical interactions, are required to have redundancy gains for the double feature targets. This can be understood as follows. For example, the C or O neurons, being single feature tuned, do not differentiate their responses to the target based on whether the target is a single feature target or a double feature target. Hence, the redundancy gain for the CO target requires that the CO cells respond more strongly to a double feature rather than a single feature target. To realize this, the suppression on the CO cells preferring and responding to the target from the neurons preferring and responding to the non-targets should be weaker when the target differs from the non-targets in two rather than one feature dimensions. This decreasing suppression by an increasing number of feature dimensions to distinguish the target can arise from three mechanisms, see Fig. 8. First, suppression between two CO cells is weaker when they prefer different features in both dimensions, rather than just one. Accordingly, suppression from a CO cell preferring and responding to non-targets on the CO cell preferring and responding to the target is weaker when the target is a double rather than a single feature target (compare Fig. 8B with Fig. 8DF). Second, suppression between a CO cell and a C cell is weaker when they prefer different colors. Accordingly, suppression from a C cell preferring and responding to a non-target to a CO cell preferring and responding to the target is weaker when the target and non-targets differ in color (compare Fig. 8BD with Fig. 8F). Third, suppression between a CO cell and an O cell is weaker when they prefer different orientations. Accordingly, suppression from a O cell preferring and responding to a non-target to a CO cell preferring and responding to the target is weaker when the target and non-targets differ in orientation (compare Fig. 8BF with Fig. 8D). The first mechanism alone should be sufficient, but either the second or third mechanism alone would not be. Future experiments, especially physiological and anatomical investigations, are needed to find out which sources are actually involved. Analogous conclusions apply to the MO cells and their associated feature tunings.

The roles of single feature and conjunctive feature tuned cells in single feature target

Our behavioral data could not reveal whether the conjunctive cells are more active than the single feature tuned cells in response to the single feature targets to dictate their saliency at least occasionally. For example, the RT_C for the C target does not reveal whether a C cell or a CO cell is responsible. After all, the value of saliency is feature blind, signaled by the firing rate of the most activated V1 neuron regardless of its preferred feature(s)(6). Our analysis has for simplicity regarded the single feature tuned cells alone as the dictating neurons for the saliencies of the single feature targets, even though the dictating responses could be from double feature tuned cells. Since these dictating responses to the single feature targets are used as the basis to calculate the contributions by the single feature tuned cells to the saliency of a double feature target, these contributions (e.g., Γ_C and Γ_O) may be over-estimating the actual contributions by the single feature tuned neurons.

Suppressions on the neurons responding to the red–vertical target
on the single feature tuned cells   on the double feature tuned cell

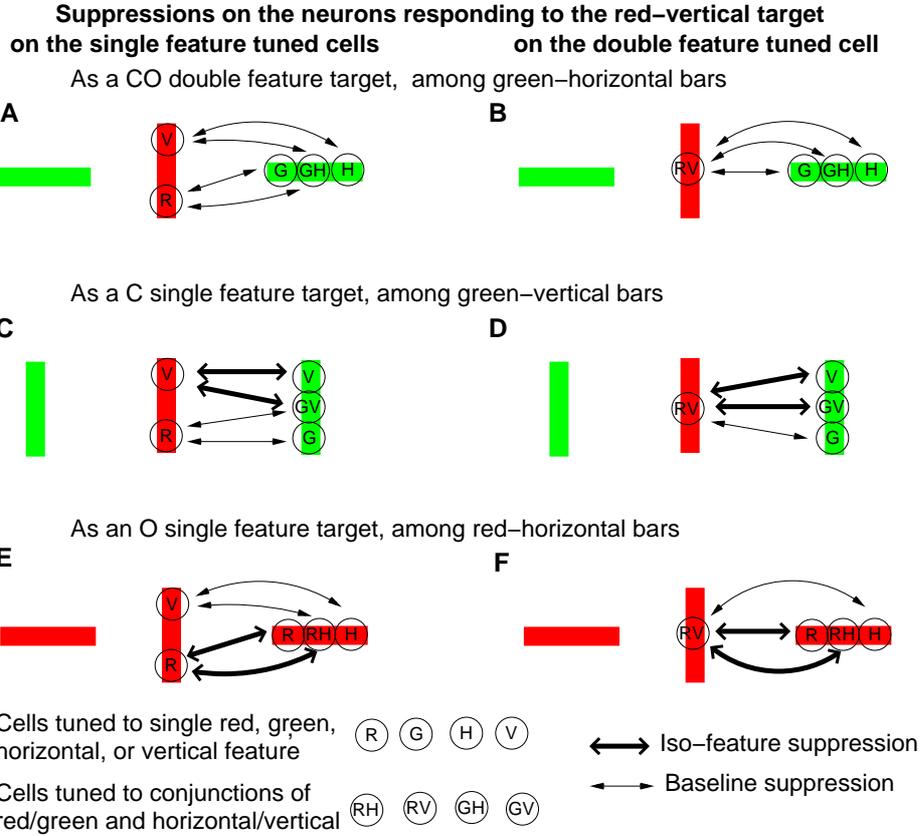As a CO double feature target, among green–horizontal bars

Figure 8: A schematic for suppression on neurons responding to the target in feature singleton search. Cases for a CO target (A and B), a C target (C and D), and a O target (E and F) are shown separately. Red and green bars are visual inputs. Circles on a bar mark neurons activated by the bar. Each neuron is marked by its preferred feature as red (R), green (G), horizontal (H), vertical (V), or a conjunction of them. Lines and curves with arrows mark (effective) suppression between two neurons, thicker for stronger suppression when the two neurons prefer the same feature. For clarity, suppression on the single feature tuned cells are shown separately (in A, C, and E) from that (in B, D, and F) on the double feature tuned cells, and interactions between the neurons responding to non-targets are not shown. Among single feature tuned neurons activated by the target, the C neuron ('R') is more suppressed when the target is an O target, whereas the O neuron ('V') is more suppressed when the target is a C target. Without the conjunctive neurons, the strongest response evoked by the CO target will be the same as the larger one of the strongest responses evoked by the C and the O targets. The CO neuron ('RV') responding to the target is least suppressed for the CO target (to have the redundancy gain), if suppression between conjunctive neurons is weaker when their preferred features are different in both rather than one feature dimensions (i.e., $W_{11} < W_{10}$ and $W_{11} < W_{01}$, see equation (19)), or if the suppression from a single feature tuned neuron (C or O neuron) on a conjunctive neuron is weaker when they prefer different features in their shared feature dimension.

21

Consequently, contributions by the double feature cells to the saliencies of the double feature targets may be under-estimated. In other words, our reported contributions Γ_{CO} and Γ_{MO} (and even Γ_{CM}) by the double feature tuned cells to the saliencies of double feature targets are in fact additional contributions by these cells beyond their hidden contributions not revealed by our RT data. These hidden contributions correspond to the contributions of the double feature tuned cells in the single feature targets. For example if the CO cells dictated the saliency of a C target in 25% of the trials and the saliency in an O target in 10% of the trials, the hidden contribution by the CO cells to the CO target could be about $0.25\Gamma_C + 0.1\Gamma_O$ (although the actual quantity depends on more specific details), making the total contribution $\Gamma_{CO} + 0.25\Gamma_C + 0.1\Gamma_O$ by the CO cells to the CO target. Analogous arguments apply to the contribution by the MO cells to the MO targets. Accordingly, considering that $\Gamma_{MO} \sim \Gamma_O, \Gamma_M$, we can conclude that the dictating neuron is no less likely, and perhaps more likely, to be an MO cell than an M or an O cell.

One may ask whether the hidden contributions by the conjunctive neurons could be so much that conjunctive neurons alone dictate the saliencies of both the single and double feature targets, as if the single feature tuned neurons are invisible or absent for saliency. To answer this question, let us denote the (effective synaptic connection mediating) intra-cortical suppression between two conjunctive cells by W_{ij}, which depends on the two binary subscripts i and j for the two feature dimensions in which the neurons are tuned. Each subscript takes value 0 or 1 if the two neurons prefer the same or different features, respectively, in the corresponding feature dimension. The strongest suppression between the two conjunctive neurons is W_{00}, when the preferred features are the same in both feature dimensions. For example, two CO neurons suppress each other most when they prefer the same color *and* the same orientation. The second strongest level of suppression includes W_{01} and W_{10}, when the preferred features are different in only one feature dimension, e.g., when two CO neurons prefer different colors but the same orientation (or the same color but different orientations). For better intuition, we may refer to W_{00} as iso-double-feature suppression and W_{01} and W_{10} as iso-single-feature suppression (see Fig. 8). The weakest suppression is W_{11}, between two conjunctive neurons preferring different features in both dimensions, e.g., when the two CO neurons prefer different colors and different orientations. Feature tuning in intra-cortical suppression means that

$$W_{00} > W_{11}, \quad \text{and}, \quad W_{00} \geq W_{01}, W_{10} \geq W_{11}. \tag{18}$$

Suppression W_{10} or W_{01} is between conjunctive neurons preferring a single feature target and those preferring the non-targets (Fig. 8DF); suppression W_{11} is between conjunctive neurons preferring a double feature target and those preferring the non-targets (Fig. 8B); whereas suppression W_{00} is between conjunctive neurons preferring the non-targets (not shown in Fig. 8 to avoid clutter). We have concluded above that

$$W_{10}, W_{01} > W_{11} \text{ sufficiently} \tag{19}$$

helps to realize redundancy gains. Now, if conjunctive neurons alone have to dictate the saliencies

of the single feature targets, then

$$W_{00} > W_{10}, W_{01} \text{ sufficiently} \tag{20}$$

is necessary to make suppression stronger on the responses to the non-targets than the target. To make all feature singletons salient and to have redundancy gains in double feature targets CO and MO but not in CM targets, no C, M, O, and CM neurons are necessary in principle, provided that equations (19) and (20) hold for both the CO and MO cells. Physiologically, there are likely a whole spectrum of single and double feature selectivities in V1(48).

Relationship with other studies

V1 physiology

The current results confirmed the previous finding by Koene and Zhaoping(10) that a statistical facilitation between the RTs for the single feature targets is sufficient to account for the shorter RTs for the CM target, but not for the CO and MO targets. Findings by both studies are consistent with physiological observations that some V1 cells are tuned conjunctively to both color and orientation(21; 22; 23), others to both orientation and motion direction(38; 39), and that few cells are tuned to both color and motion direction(24). However, the current study differs from Koene and Zhaoping(10) in research questions asked, methodology, and outcomes. Koene and Zhaoping(10) used the behavioral RT data and the known V1 facts (the presence and absence of certain conjunctive cells) to test whether the V1 saliency hypothesis is correct, whereas the current study applies the V1 saliency hypothesis (assumed to be correct) to the behavioral data to investigate the less known V1 neural properties — the response levels and feature selectivities associated with the conjunctively cells. Koene and Zhaoping(10) found out qualitatively whether the RT redundancy gains are present in certain feature dimensions in order to determine whether the neural substrate for saliency is most likely V1 rather than the extrastriate cortices. In contrast, the current study formulates an optimization approach to predict quantitatively the probability that the conjunctive neurons should dominate the V1 responses. In addition, the current study predicts the feature tuning properties of the intra-cortical interactions associated with the conjunctive neurons. These predicted properties, in particular that the MO cells are no less likely than either of the single feature tuned neurons to dominate the responses to a MO singleton, can be experimentally tested.

Iso-feature suppression is only one of the intra-cortical interaction between V1 neurons, albeit a dominant one. Another notable interaction is colinear facilitation(49), the excitation between two V1 neurons whose preferred input bars have similar orientations and are aligned as if belonging to a single smooth curve. When a central bar (such as our visual search target) is surrounded by uniformly oriented background bars in a statistically isotropic manner, the net interaction between the (neurons responding to the) central bar and the surrounding bars is still iso-orientation suppression, stronger when the orientations of the central and surrounding bars are more similar, as observed physiologically(42). Since the density of our non-targets is quite high, each non-target can

be approximately viewed as surrounded by other non-targets isotropically and experiencing a net iso-orientation suppression as well. Hence, for our current study when it is only necessary to evaluate the net suppression on each neuron, it is not necessary to consider colinear facilitation separately.

The role of the extrastriate cortices for attentional guidance

There are many neurons tuned to CM conjunctions in V2(25; 26), but few in V1(24). Hence, our finding of no contribution by the CM neurons provides a strong support that V1 rather than extrastriate mechanisms play the dominant role in saliency for these feature singletons. This however does not rule out the possibility that the extrastriate cortex plays a role guiding attention for other visual stimuli. Recently, depth cue was found(33) to speed up the task to locate a texture border only if this border is not salient enough for observers to report its location within one second. Since extrastriate cortices rather than V1 are thought as responsible for depth perception(27; 28; 29; 30; 31; 32), this finding suggests that, when the target saliency is too weak, V1 signals may be insufficient to guide attention in a dominant manner. The brain areas such as the superior colliculus may coordinate and combine contributions from various cortical areas to guide attention. Superior colliculus is particularly suitable for such a role since it receives inputs from multiple brain areas including V1, extrastriate cortex, and parietal cortices, and directly controls the gaze shifts through the brain stem(50; 51). Since longer latencies are typically required for contributions from higher brain areas, it is conceivable that the speeded or hurried decisions for attentional shifts are reached using only contributions from lower cortical areas such as V1. Meanwhile, since human gaze shifts about three times per second, and since previous works suggest that top-down factors play an increasingly dominant role to guide attention when longer latencies are allowed(4; 52), it is unclear whether the attentional guidance 800 ms after the stimulus onset could be viewed as strictly by (bottom-up) saliency alone.

Implications on conjunction search

The visual search task considered in this study is a feature search task(2), since the target can be distinguished by a unique feature, even when it is a double feature target. In contrast, when a target shares one (or more) features with some or all distractors and can only be distinguished by a particular conjunction of features, the search is much more difficult and is called a conjunction search(2). For example, to find a red-vertical target bar among red-horizontal bars and green-vertical bars, the target conjunction is of red color feature and the vertical orientation feature, while both red and vertical features are present among the non-targets. The difficulty in conjunction searches can be easily understood if there is no conjunctive neurons. For the example above, the neurons preferring red respond to both the target and many non-targets and suppress each other by iso-color suppression; similarly, the neurons preferring vertical respond to both the target and many non-targets and suppress each other by iso-orientation suppression (see the left half of Fig. 9). Consequently, the single feature tuned neurons cannot distinguish a target by their response levels since their responses to the target are no higher statistically than their responses to the non-targets.
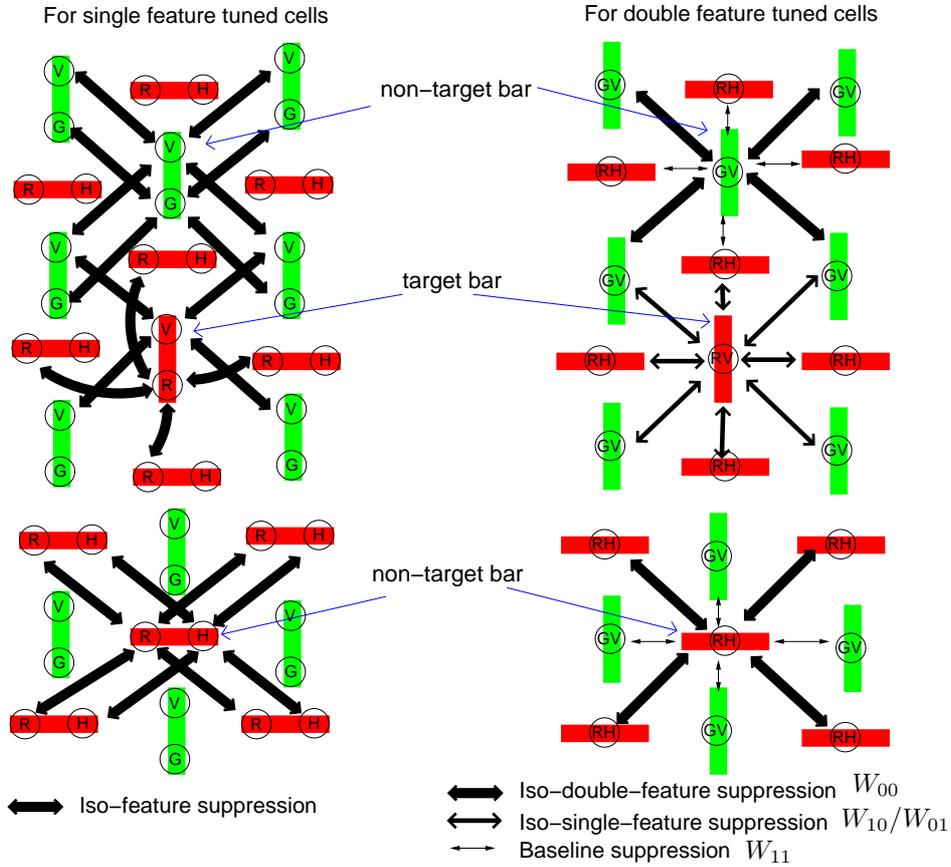
Figure 9: A schematic for neurons and their interactions in a conjunction search for a red-vertical target. Bars and neurons are similarly visualized as in Fig. 8. For clarity, interactions between single feature tuned neurons are shown separately (on the left) from those between conjunctive neurons. To avoid clutter, only interactions associated with neurons responding to the target bar and two of the non-target bars are shown, the baseline suppression on the single feature tuned cells and suppression between single feature tuned and conjunctive neurons are not shown. Each single feature tuned neuron, regardless of its preferred feature and regardless of whether it is responding to the target, experiences iso-feature suppression from neurons responding to about half of the neighboring bars. Hence, single feature tuned neurons cannot distinguish the target by their response levels. The conjunctive neuron ('RV') responding to the target experiences iso-single-feature suppression (W_{01} or W_{10}) from other conjunctive neurons responding to all neighboring bars, whereas each conjunctive neuron responding to a non-target experiences iso-double-feature suppression (W_{00}) from conjunctive neurons responding to only half of the neighboring bars. The target cannot be distinguished by a higher response (from the 'RV' neuron) if $W_{01} + W_{10} \gtrsim W_{00} + W_{11}$, or if this neuron's response is weaker than the responses from the single feature tuned neurons.

25

However, one may wonder whether the conjunctive neurons preferring the unique target conjunction could distinguish the target by a relatively higher response. After all, the conjunctive neurons could respond as vigorously as the single feature tuned neurons to double feature singletons, and the conjunctive cell preferring and responding to the target may largely escape the suppression from neurons preferring and responding to the non-target conjunctions.

This question can be answered by dissecting the intra-cortical interactions associated with the conjunctive neurons only, see the right half of Fig. 9. We again use the example of the conjunctive search for red-vertical, and take for simplicity the (most difficult) situation when half of the non-targets are green-vertical and the other half are red-horizontal. Each conjunctive neuron preferring and responding to a non-target item (e.g., green-vertical) is subject to strong iso-double-feature suppression W_{00} from other conjunctive neurons preferring the same color *and* the same orientation and responding to half of the non-targets in its vicinity. It should largely escape iso-feature suppression, or experience a much weaker suppression W_{11}, from the conjunctive neurons preferring and responding to the other half of the non-targets (e.g., red-horizontal) in the vicinity, since they prefer different color *and* different orientation. Meanwhile, a neuron preferring red-vertical and responding to the target is subject to two sources of iso-single-feature suppression: iso-color suppression W_{01} from red-horizontal preferring neurons responding to half of the non-targets in the vicinity, and iso-orientation suppression W_{10} from green-vertical preferring neurons responding to the other half of the non-targets in the vicinity. Hence, the response to the target is subject to suppression W_{01} or W_{10} from neurons responding to all neighboring items, whereas the response to a non-target is subject to suppression W_{00} from neurons responding to only (about) half of the neighbors. Therefore, the response to the target is not distinguished unless $W_{01} + W_{10}$ is sufficiently weaker than W_{00} (or $W_{00} + W_{11}$ when including suppression from the conjunctive neurons preferring different features in both dimensions). For example, if two conjunctive cells do not substantially suppress each other unless they prefer the same feature in both feature dimensions, i.e., $W_{00} \gg W_{10} \approx 0$, $W_{00} \gg W_{01} \approx 0$, then the response to the unique target conjunction can be relatively free of suppression to make the target salient. This situation has been demonstrated in a V1 model, see Fig. 5 of Li(6).

It should be noted that the arguments above have for simplicity omitted the interactions between the single feature tuned cells and the double feature tuned cells. It is also possible that the conjunctive cell responding to the target is suppressed by the single feature tuned cells responding to the neighboring non-targets, since each non-target shares one feature in common with the target. The same conclusion in the last paragraph could still be reached if the iso-single-feature suppression (W_{10} and W_{01}) between two conjunctive cells is replaced by the iso-feature suppression between a conjunctive neuron and a single feature tuned neuron preferring the same feature in their shared feature dimension. As far as a conjunctive neuron is concerned, the pre-synaptic source for the iso-single-feature suppression may be either the double or single feature tuned neurons, or may include both. Similarly, iso-feature suppression on single feature tuned cells could arise from both the single feature tuned and conjunctively tuned cells.

Since color-orientation conjunction search is known to be difficult(2), it suggests that iso-feature

contextual suppression on a CO cell (responding to its preferred input) is substantial even when the contextual inputs is different from the preferred input in one, but not both, of the two feature dimensions. This conclusion, also reached previously(10; 11), can be physiologically tested. Meanwhile, McLeod, Driver, and Crisp(53) showed that the conjunction search for a moving "X" among static "X"s and moving "O"s are relatively easy. If one treats the difference between an "X" and an "O" as a difference in orientation, this suggests that the a MO conjunction search is not too difficult, and if so, one could infer that MO neurons are not sufficiently suppressed by contextual inputs unless the contextual inputs and the preferred inputs share the same feature in both the O and M dimensions. However, a more authentic MO conjunction search is required for more confident inferences.

It is now clear that a conjunctive search should definitely be difficult if there is no V1 neurons preferring this particular conjunction. For example, there is no V1 neuron which simultaneously prefers two different orientations (or two colors) without also preferring the average orientation (or color) of the two preferred ones. Indeed, it has long been known that a unique conjunction of two different features within a single feature dimension, e.g., a conjunction of two orientations, is very difficult to find(54). Similarly, redundancy gains involving two features in the same dimension should be absent, consistent with behavioral data(11).

Redundancy gains for saliency versus feature integration for object recognition

It takes longer to identify both features, color and orientation, of an object than it is to identify just one feature(55). This is in contrast to the shorter reaction times to find a feature singleton unique in two, rather than one, feature dimensions in visual search. These two situations involve two different tasks, one is object recognition or identification and the other is feature detection or localization. These two tasks are often called the "what" and "where" tasks, respectively, and are believed to involve separate brain regions(56). By the psychological Feature Integration Theory(2), additional processing is needed to bind two features of a single object to identify the object after the location of the object is selected by spatial attention. Meanwhile, the feature singleton detection in our task mainly involves bottom-up saliency to select the most salient location without identifying the features or objects. Indeed, observers for the task typically did not pay attention to which features distinguish the target when they pressed the button to report its location(10). A separation between the "where" and "what" task is one of the foundations of the theoretical framework that V1 mechanisms serve the functional role of visual segmentation without classification(40), which means to segment an image region (by highlighting its boundaries with higher V1 responses) without recognizing the region. This theoretical framework has in turn inspired the V1 saliency hypothesis, which uses V1 activities to represent saliencies before decoding the visual input feature values from the very same activities(6). Accordingly, the V1 neural activities are universal currencies for saliency regardless of their feature preferences(57).

Concluding remarks

The V1 saliency hypothesis enables us to probe the properties of V1 neurons and intra-cortical interactions from behavioral data on visual search tasks, rather than by physiological experiments. Inferring coarse scale brain substrates from behavior is quite common in psychological studies. For example, damage to hippocampus could be inferred if somebody has difficulty forming new memories, applying the knowledge that hippocampus is a substrate for memory consolidation(58). However, inferring neuronal level details from behavior is much less common. Many of the previous works linking physiology and behavior are to explain behavior from physiology. For example, sensory discrimination thresholds can be derived from feature tuning of the neurons and the densities of neurons involved(59; 60; 61; 62). Works to infer physiology from behavior are mainly those to infer the underlying neural channels of signal representation via sensory adaptation(63). Behind these works are theories of optimal sensory decoding or assumptions linking neural sensitivities to behavioral sensitivities. The current work adds the V1 saliency hypothesis to the theoretical bases that can be used to link physiology to behavior, thereby extending the realms of neural mechanisms that can be probed from behavior.

References

[1] Itti L, Koch C (2001) Computational modeling of visual attention. Nature Reviews Neuroscience 2: 194-203.

[2] Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognitive Psychology 12: 97-136.

[3] Posner MI, Petersen SE (1990) The attention system of the human brain. Annual Review of Neuroscience 13: 25-42.

[4] Nakayama K, Mackeben M (1989) Sustained and transient components of focal visual attention. Vision research 29: 1631-1647.

[5] Li Z (1999) Contextual influences in V1 as a basis for pop out and asymmetry in visual search. Proc Natl Acad Sci USA, 96(18): 10530-10535.

[6] Li Z (2002) A saliency map in primary visual cortex. Trends in Cognitive Sciences 6: 9-16.

[7] Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. Psychological review 96: 433-458.

[8] Wolfe JM, Cave KR, Franzel SL (1989) Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human perception and performance 15: 419-433.

[9] Zhaoping L, Snowden RJ (2006) A theory of a saliency map in primary visual cortex (V1) tested by psychophysics of colourorientation interference in texture segmentation. Visual Cognition 14(4/5/6/7/8): 911-933.

[10] Koene AR, Zhaoping L (2007) Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. Journal of Vision 7(7): 6, 1-14.

[11] Zhaoping L, May KA (2007) Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. PLoS Computational Biology 3(4): e62.

[12] Jingling L, Zhaoping L (2008) Change detection is easier at texture border bars when they are parallel to the border: Evidence for V1 mechanisms of bottom-up salience. Perception 37(2): 197-206.

[13] Zhaoping L (2008) Attention capture by eye of origin singletons even without awareness–a hallmark of a bottom-up saliency map in the primary visual cortex. Journal of Vision 8(5): 1, 1-18.

[14] Zhang X, Zhaoping L, Zhou T, Fang F (2012) Neural activities in V1 create a bottom-up saliency map. Neuron 73(1): 183-192.

[15] Zhaoping L (2012) Gaze capture by eye-of-origin singletons: Interdependence with awareness. Journal of Vision 12(2): 17,1-22.

[16] Raab DH (1962) Statistical facilitation of simple reaction times. Transactions of the New York Academy of Sciences 24: 574-590.

[17] Miller J (1978) Multidimensional same–different judgments: Evidence against independent comparisons of dimensions. Journal of Experimental Psychology: Human perception and performance 4: 411-422.

[18] Miller J (1982) Divided attention: Evidence for coactivation with redundant signals. Cognitive psychology 14: 247-279.

[19] Ulrich R, Giray M (1986) Separate-activation models with variable base times: Testability and checking of cross-channel dependency. Attention, Perception, & Psychophysics 39: 248-254.

[20] Krummenacher J, Müller HJ, Heller D (2002) Visual search for dimensionally redundant pop-out targets: Parallel-coactive processing of dimensions is location specific. Journal of Experimental Psychology: Human Perception and Performance 28: 1303-1322.

[21] Livingstone MS, Hubel DH (1984) Anatomy and physiology of a color system in the primate visual cortex. The Journal of Neuroscience 4: 309-356.

[22] Ts'o DY, Gilbert CD (1988) The organization of chromatic and spatial interactions in the primate striate cortex. The Journal of Neuroscience 8: 1712-1727.

[23] Friedman HS, Zhou H, von der Heydt R (2003) The coding of uniform colour figures in monkey visual cortex. The Journal of physiology 548: 593-613.

[24] Horwitz GD, Albright TD (2005) Paucity of chromatic linear motion detectors in macaque V1. Journal of Vision 5(6): 4, 525-533.

[25] Gegenfurtner KR, Kiper DC, Fenstemaker SB (1996) Processing of color, form, and motion in macaque area V2. Visual neuroscience 13: 161-172.

[26] Shipp S, Adams DL, Moutoussis K, Zeki S (2009) Feature binding in the feedback layers of area V2. Cerebral Cortex 19: 2230-2239.

[27] DeAngelis GC, Cumming BG, Newsome WT (1998) Cortical area mt and the perception of stereoscopic depth. Nature 394: 677-680.

[28] Bakin JS, Nakayama K, Gilbert CD (2000) Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. The Journal of Neuroscience 20: 8188-8198.

[29] Cumming BG, Parker AJ (2000) Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. The Journal of Neuroscience 20: 4758-4767.

[30] von der Heydt R, Zhou H, Friedman HS (2000) Representation of stereoscopic edges in monkey visual cortex. Vision Research 40: 1955-1967.

[31] Janssen P, Vogels R, Liu Y, Orban GA (2003) At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. Neuron 37: 693-701.

[32] Qiu FT, von der Heydt R (2005) Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. Neuron 47: 155-166.

[33] Zhaoping L, Guyader N, Lewis A (2009) Relative contributions of 2D and 3D cues in a texture segmentation task, implications for the roles of striate and extrastriate cortex in attentional selection. Journal of Vision 9(11): 20,1-22.

[34] van Zoest W, Donk M (2006) Saccadic target selection as a function of time. Spatial Vision 19: 61-76.

[35] Lamme VA, Supèr H, Spekreijse H (1998) Feedforward, horizontal, and feedback processing in the visual cortex. Current Opinion in Neurobiology 8: 529-535.

[36] Girard P, Hupé JM, Bullier J (2001) Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. Journal of Neurophysiology 85: 1328-1331.

[37] Angelucci A, Levitt JB, Walton EJS, Hupé JM, Bullier J, et al. (2002) Circuits for local and global signal integration in primary visual cortex. The Journal of Neuroscience 22: 8633-8646.

[38] Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. The Journal of Physiology 148: 574-591.

[39] Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. The Journal of Physiology 195: 215-243.

[40] Li Z (1999) Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. Network: Comput Neural Syst 10: 187-212.

[41] Allman J, Miezin F, McGuinness E (1985) Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (mt). Perception 14: 105-126.

[42] Knierim JJ, van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. Journal of Neurophysiology 67: 961-980.

[43] Li CY, Li W (1994) Extensive integration field beyond the classical receptive field of cat's striate cortical neurons–classification and tuning properties. Vision research 34: 2337-2355.

[44] Wachtler T, Sejnowski TJ, Albright TD (2003) Representation of color stimuli in awake macaque primary visual cortex. Neuron 37: 681-691.

[45] Jones HE, Grieve KL, Wang W, Sillito AM (2001) Surround suppression in primate V1. J Neurophysiol 86: 2011-2028.

[46] Li Z (2001) Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. Neural Computation 13(8): 1749-1780.

[47] Zhaoping L (2003) V1 mechanisms and some figure-ground and border effects. Journal of Physiology, Paris 97(4-6): 503-515.

[48] Lennie P (1998) Single units and visual cortical organization. Perception 27: 889-935.

[49] Kapadia M, Ito M, Gilbert C, Westheimer G (1995) Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron 15: 843-56.

[50] Schiller P (1998) The neural control of visually guided eye movements. In: Richards JE, editor, Cognitive Neuroscience of Attention, a developmental perspective, Mahwah, New Jersey USA.: Lawrence Erlbaum Associates, Inc. pp. 3-50.

[51] Shipp S (2004) The brain circuitry of attention. Trends in Cognitive Sciences 8: 223-230.

[52] van Zoest W, Donk M, Theeuwes J (2004) The role of stimulus-driven and goal-driven control in saccadic visual selection. Journal of Experimental Psychology: Human perception and performance 30: 746-759.

[53] McLeod P, Driver J, Crisp J (1988) Visual search for a conjunction of movement and form is parallel. Nature 332: 154-155.

[54] Wolfe J, Yu K, Stewart M, Shorter A, Friedman-Hill S, et al. (1990) Limitations on the parallel guidance of visual search: Color× color and orientation× orientation conjunctions. Journal of Experimental Psychology: Human Perception and Performance 16: 879-892.

[55] Bodelon C, Fallah M, Reynolds J (2007) Temporal resolution for the perception of features and conjunctions. The Journal of Neuroscience 27: 725-730.

[56] Mishkin M, Ungerleider L, Macko K (1983) Object vision and spatial vision: Two cortical pathways. Trends in neurosciences 6: 414–417.

[57] Zhaoping L (2006) Theoretical understanding of the early visual processes by data compression and data selection. Network: computation in neural systems 17(4): 301-334.

[58] Squire L (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychological review 99: 195-231.

[59] Moiseff A, Konishi M (1981) Neuronal and behavioral sensitivity to binaural time differences in the owl. The Journal of Neuroscience 1: 40-48.

[60] Paradiso MA (1988) A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. Biological Cybernetics 58: 35-49.

[61] Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: A comparison of neuronal and psychophysical performance. The Journal of Neuroscience 12: 4745-4765.

[62] Zhaoping L, Geisler WS, May KA (2011) Human wavelength discrimination of monochromatic light explained by optimal wavelength decoding of light of unknown intensity. PloS One 6(5): e19248.

[63] Blakemore C, Campbell FW (1969) On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. The Journal of Physiology 203: 237-260.