Behavioral/Systems/Cognitive

Flexible Categorization of Relative Stimulus Strength by the Optic Tectum

Shreesh P. Mysore and Eric I. Knudsen
Department of Neurobiology, Stanford University, Stanford, California 94305

Categorization is the process by which the brain segregates continuously variable stimuli into discrete groups. We report that patterns of neural population activity in the owl optic tectum (OT) categorize stimuli based on their relative strengths into "strongest" versus "other." The category boundary shifts adaptively to track changes in the absolute strength of the strongest stimulus. This population-wide categorization is mediated by the responses of a small subset of neurons. Our data constitute the first direct demonstration of explicit categorization of stimuli by a neural network based on relative stimulus strength or salience. The finding of categorization by the population code relaxes constraints on the properties of downstream decoders that might read out the location of the strongest stimulus. These results indicate that the ensemble neural code in the OT could mediate bottom-up stimulus selection for gaze and attention, a form of stimulus categorization in which the category boundary often shifts within hundreds of milliseconds.

Introduction

The ability of the brain to categorize continuously variable stimuli into discrete groups is essential to adaptive behavior. Categorization permits neural responses to be insensitive to differences among stimuli that are not relevant to behavior, and to be particularly sensitive to differences among stimuli that are. In so doing, categorization facilitates higher brain functions such as perception and decision making (Leopold and Logothetis, 1999; Wang, 2008).

Bottom-up stimulus selection, the selection of the strongest (most "salient") stimulus in the environment, is a critical component of the selection of the next target for gaze or attention (Itti and Koch, 2001). It can be thought of as a form of flexible categorization in which stimuli are categorized into two classes: "strongest" and "others" (Fig. 1*A*). However, unlike tasks in which category boundaries are fixed or learned (Freedman and Assad, 2006; Prather et al., 2009), the category boundary in bottom-up stimulus selection needs to change in real-time depending on the specific set of stimuli in the environment. Since the locus of gaze or attention can change as frequently as several times a second, the circuitry underlying this categorization must be capable of changing the category boundary flexibly and dynamically (Fig. 1*A*). Although several brain areas are critically involved in stimulus selection (e.g., lateral intraparietal area, LIP; frontal eye field, FEF; superior colliculus, SC), it is unclear how and where physical salience (stimulus strength)-dependent categorization is explicitly represented in the brain.

The intermediate and deep layers of the superior colliculus (SCid) in monkeys are known to be critically involved in competitive stimulus selection (Carello and Krauzlis, 2004; McPeek and Keller, 2004; Müller et al., 2005; Lovejoy and Krauzlis, 2010). Recently, we examined the representation of relative stimulus strength in the intermediate and deep layers of the owl OT (OTid, equivalent to the mammalian SCid). We found that for a subpopulation of neurons, responses to a stimulus inside the classical receptive field (RF) are suppressed abruptly by an increase in the strength of a competing stimulus outside the RF (Mysore et al., 2011). The responses of these "switch-like" neurons dropped from a high level when the stimulus inside the RF was the strongest stimulus to a low level when the competing stimulus became the strongest stimulus. This property was independent of the sensory modality of the competing stimulus. An important unanswered question is how such a representation might be decoded by downstream neurons. Specifically, does a decoder require access selectively to switch-like responses to reliably identify the strongest stimulus? We addressed this question by using a new strength-morphing protocol to test whether the population code in the OTid provides an explicit categorical representation of the strongest stimulus.

Materials and Methods

Neurophysiology. Experiments were performed in five untrained, nontranquilized, head-fixed owls (both male and female) following experimental protocols described previously (Mysore et al., 2010, 2011). Briefly, epoxy-coated tungsten microelectrodes (FHC, 250 μm, 1–5 MΩ at 1 kHz) were positioned in the OTid of owls tranquilized with nitrous oxide. Nitrous oxide was then turned off and single and multiple units were recorded in nontranquilized owls. Extracellularly recorded multiunit spike waveforms were sorted offline into putative single units (Fee et al., 1996; Mitra and Bokil, 2008).

Stimuli. The looming (expanding) visual stimuli used here and the responses of OTid neurons to looming stimuli have been described pre-

Received Oct. 10, 2010; revised March 31, 2011; accepted April 14, 2011.

Author contributions: S.P.M. and E.I.K. designed research; S.P.M. performed research; S.P.M. contributed unpublished reagents/analytic tools; S.P.M. analyzed data; S.P.M. and E.I.K. wrote the paper.

Funding from the Stanford University School of Medicine (Dean's Postdoctoral Fellowship, SPM) and the NIH (9 R01 EY019179-30, E.I.K.) is gratefully acknowledged. We thank Alex Goddard, Ali Asadollahi, Astra Bryant, Cathy Dunn, and Devarajan Sridharan for critically reading the manuscript.

Correspondence should be addressed to Shreesh P. Mysore, 299 W. Campus Drive, Department of Neurobiology, Stanford University, Stanford, CA 94305. E-mail: shreesh@stanford.edu.

DOI:10.1523/JNEUROSCI.5425-10.2011

viously (Mysore et al., 2010). Briefly, looming stimuli were dots that expanded linearly in size over time, starting from a size of 0.6° in radius. Loom speeds are denoted in the text as percentage of maximum where maximum speed equals 16°/s. The spatial RF of a unit was defined as the set of locations at which a single stimulus evoked responses above baseline. The average size of visual RFs was 11.8 ± 0.6° (width of the azimuthal tuning curve at half-maximum response). In comparison, the final size of the fastest looming stimulus used in most experiments was 9.2° (diameter), except for the final set of experiments (see Fig. 5), in which it was 11° (diameter). Thus, the looming stimuli were typically contained within the visual RF. Nonetheless, the final size of a looming stimulus is not the primary determinant of the strength of the response it evokes; rather, it is the speed of the loom (rate of change of dot size) per se (Mysore et al., 2010).

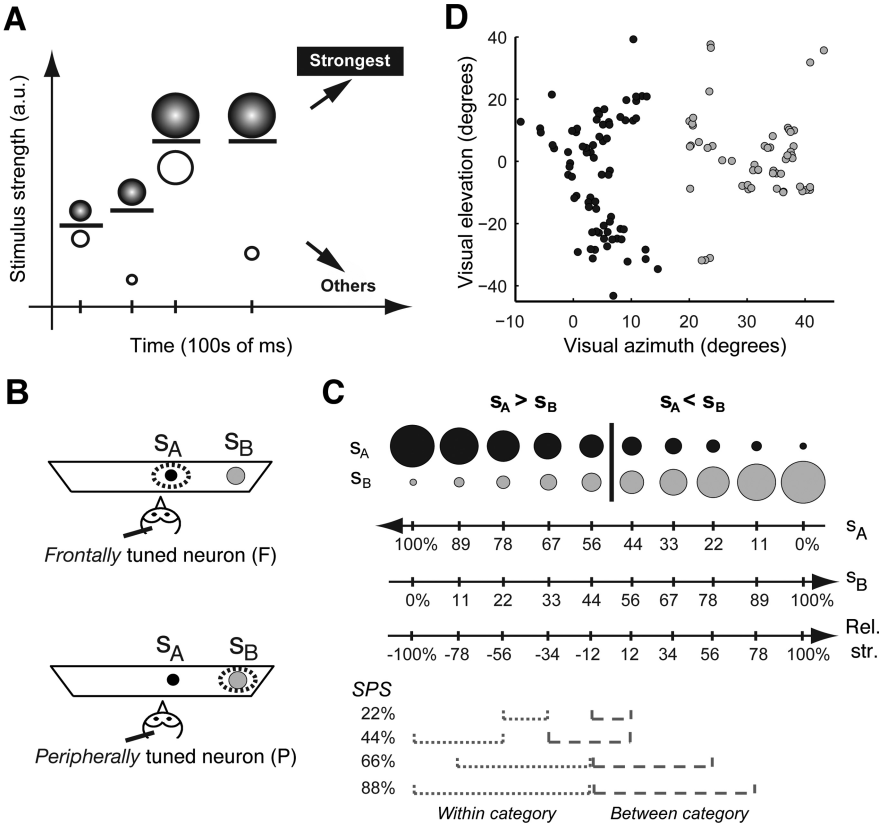

Figure 1. Stimulus selection as flexible categorization. ***A***, The stimulus pair at each instant (shaded and open dots) represents the set of stimuli from which the strongest is to be selected. Dot size denotes the strength of the stimulus; horizontal lines indicate the dynamically shifting category boundary. ***B***, Schematic representation of neuronal recordings from frontally tuned (top) and peripherally tuned (bottom) OTid neurons in passive, untrained owls. Shown are the owl, electrode, tangent screen, visual RFs (dotted areas), and the two looming visual stimuli s_A and s_B. In experiments, both stimuli were full-contrast black dots on a light background; they are represented here in different shades of gray for clarity. ***C***, Schematic representation of the 10 different pairs of stimulus strengths (s_A/s_B) used in the strength-morphing protocol (Materials and Methods). Filled black circles, s_A. Filled gray circles, s_B. Dot size represents strength of the stimulus. The two categories of stimulus strength ($s_A > s_B$ and $s_A < s_B$) are indicated. *x*-axis labels indicate, in order, the strength of s_A, the strength of s_B, and the relative strength between s_A and s_B ($s_B - s_A$) for each strength pair. Strengths represented as percentage of the maximum loom speed (16°/s; Materials and Methods). Indicated below the *x*-axis labels are examples of within-category and between-category SPSs used in quantifying the strength of categorization (Materials and Methods). ***D***, Visual RF centers of frontally (filled black circles) and peripherally (filled gray circles) tuned OTid neurons recorded in this study.

Strength-morphing protocol. Two visual stimuli, s_A and s_B, were presented simultaneously such that one was centered inside the RF of the recorded OTid neuron while the other was presented outside the RF (30° away in azimuth from the RF center) (Fig. 1*B*). Stimulus s_A was always presented frontally (<20° azimuth), while stimulus s_B was always presented peripherally (≥20° azimuth). Positive values of azimuth denote locations contralateral to the recording site.

Neurons in the OTid are known to respond increasingly strongly to increasing strengths of stimuli (faster speeds, higher contrasts, higher amplitude sounds, etc). In our experimental protocol, we systematically "morphed" the relative strengths of the two stimuli: as the strength of one stimulus (s_A) decreased, the strength of the other (s_B) increased (Fig. 1*C*). Both s_A and s_B were looming dots, and their loom speeds corresponded to their strengths. The 10 strength pairs s_A/s_B used in our protocol corresponded to 100%/0%, 89%/11%, 78%/22%, 67%/33%, 56%/44%, 44%/56%, 33%/67%, 22%/78%, 11%/89%, 0%/100%, with percentages calculated relative to a loom speed of 16°/s; note that a stimulus with 0% loom speed was a stationary dot. The 10 strength pairs are schematized in Figure 1*C* as vertical pairs of black and gray dots. The relative strength corresponding to each strength pair is indicated along the *x*-axis, with positive values indicating a stronger s_B stimulus. Because the range of 100% cannot be sampled at 10 equally spaced integer values, we chose samples such that the interval between the two central samples (44% and 56%) was 12%, and the intervals between all other pairs of neighboring samples were 11% (Fig. 1*C*). For analysis, however, all intervals were treated as being equal.

In contrast to the protocol that we used previously to vary relative stimulus strength (Mysore et al., 2011), in which the strength of the stimulus inside the RF remained fixed while that of the stimulus outside the RF was varied, the strength-morphing protocol used here afforded the advantage that relative strength was varied while the average drive to the network remained constant.

Criteria for including units in sample. Units were first tested for visual responsiveness. The visual RFs of responsive units were mapped out. Units were then tested for sensitivity to loom speeds by examining whether their responses were positively correlated with the speed of a looming stimulus centered in the RF. The range of loom speeds tested was within 0–22°/s [shown to be the range that most OTid units encode with increasing firing rates (Mysore et al., 2010, 2011)]. Units that were sensitive to loom speeds were further tested with the strength-morphing protocol. Of all the units studied in this manner, those for which the maximum firing rate response was below the 95th percentile value for the population were included in the estimates of the patterns of population responses (64/68 for the analysis in Fig. 2, and 49/51 for the analysis in Fig. 5).

Data analysis. The stimulus categories we tested were $s_A > s_B$ and $s_A < s_B$. All analyses were performed with custom MATLAB code. The responses of single neurons to the series of s_A/s_B strength pairs were measured by counting spikes over a 100–250 ms window with respect to stimulus onset, and converting the resulting counts into spikes per second. To quantify the degree to which these responses categorized $s_A > s_B$ and $s_A < s_B$, we compared responses of a population of OTid neurons (population response pattern) to the different strength pairs by applying techniques used previously to study categorization of odors in the olfactory bulb and of visual motion direction in the parietal cortex (Freedman and Assad, 2006; Niessing and Friedrich, 2010). Population response patterns were compared, by correlation analysis, across different combinations of s_A/s_B strength pairs. The correlation values (Pearson's correlation coefficient; corr command in MATLAB) were organized in a matrix, the rows (and columns) of which represented the different strength pairs. The degree to which the population response pattern for one strength pair correlated with the population response patterns for all other strength pairs appears as vertical and horizontal transects through the matrix.

To test for stimulus categorization, we compared two metrics: (1) the average within-category difference (WCD) in response correlations for strength pairs in the same category ($s_A > s_B$ or $s_A < s_{B)}$, and (2) the average between-category difference (BCD) in response correlations for strength pairs in different categories. For this analysis, the separation in the relative strengths of any two strength pairs was designated as the "strength pair separation" (SPS) (Fig. 1*C*). Examples of strength pair separations for both the within-category and between-category calculations are shown at the bottom of Figure 1*C*. The topmost within-category separation of 22% is shown for strength pairs with relative strengths −56% and −34% (Fig. 1*C*, dotted bracket on left). Seven other strength pairs also had a separation of 22%, for example, strength pairs with relative strengths of −100% and −78% or 12% and 34%, etc (the separations indicated in Fig. 1*C* were chosen randomly). Most strength pair separations could be achieved using different combinations of strength pairs. For within-category comparisons, the maximum numbers of strength pair combinations for achieving a given strength pair separation were as follows: 4 × 2 for 22% separation (×2 reflects the two categories), 3 × 2 for 44%, 2 × 2 for 66%, and 1 × 2 for 88%. For between-category comparisons, the maximum numbers of combinations were as follows: 1 for a 22% separation, 2 for 44%, 3 for 66%, and 4 for 88%.

For the comparison of WCD and BCD to be valid, it was necessary to ensure that the following two constraints were satisfied: that both calculations included equal numbers of strength pair separations, and that only those separations were included that were represented in both WCD and BCD calculations. To satisfy these two constraints, both WCD and BCD values were calculated using eight strength pair separations: 22% (1 combination), 44% (2 combinations), 66% (3 combinations), and 88% (2 combinations). Strength pair separations of >88% (such as 110%) were not included in the BCD calculation because no such separations could be included in the WCD calculation (including such separations in the BCD calculation would bias the BCD toward larger values).

The error bars (SEM values) for the WCD and BCD calculations (Figs. 2*E*, 4*E*, 5*C*) were estimated using a bootstrap procedure (3000 resamplings) in which combinations of strength pairs that yielded the same strength pair separation were chosen randomly from the maximum possible number of combinations for that separation. For instance, for a within-category separation of 44%, the two allowed combinations were chosen randomly from the six (3 × 2) possible combinations.

The WCD and BCD values in Figure 5*C* were calculated using a similar procedure as above with the modification that nine strength pair separations were chosen (instead of eight). This change reflected the fact that in Figure 5, one category ($s_A > s_B$) contained six strength pairs and the other contained four, instead of five in each category as in Figure 2. The nine included strength pair separations were as follows: 22% (1 combination), 44% (2 combinations), 66% (3 combinations), 88% (2 combinations), and 110% (1 combination).

The strength of categorization was quantified using a categorization index that summarized the relationship between the WCD and BCD values. It was defined as (BCD − WCD)/(BCD + WCD). Positive values of the index indicate smaller differences within a category, while negative values indicate larger differences within a category.

Results

We examined stimulus strength-dependent categorization in the OTid of passive, untrained owls using pairs of looming dots in a strength-morphing protocol (Materials and Methods and Fig. 1*C*). In this protocol, stimulus s_A was always presented frontally (<20° azimuth), and stimulus s_B was always presented peripherally, 30° away. Neurons were recorded such that either s_A was centered in the RF (frontally tuned neurons) (Fig. 1*B*, top; and Fig. 1*D*, black), or s_B was centered in the RF (peripherally tuned neurons) (Fig. 1*B*, bottom; and Fig. 1*D*, gray). The combined responses of frontally tuned neurons provided an estimate of the population response to stimulus s_A, while those of peripherally tuned neurons provided an estimate of the population response to stimulus s_B. Relative stimulus strength in this protocol was varied by increasing the strength of s_B while simultaneously decreasing the strength of s_A (Fig. 1*C*). All strength pairs were presented in a randomly interleaved manner.

Population representation of relative stimulus strength is categorical

Responses of a frontally and a peripherally tuned neuron to the strength-morphing protocol are shown in Figure 2*A*. As expected, as the strength of s_B increased (and that of s_A simultaneously decreased), the responses of the peripherally tuned neuron increased (Fig. 2*A*, blue), while those of the frontally tuned neuron decreased (Fig. 2*A*, red).

To infer the pattern of responses across the population of OTid neurons to the strength-morphing protocol, we adopted the approach used by Niessing and Friedrich (2010). We combined the responses of both frontally and peripherally tuned neurons into a matrix (rows represented neurons and columns represented strength pairs) (Fig. 2*B*) (Materials and Methods). Each column of this matrix represented an estimate of the pattern of responses across the OTid population to a particular s_A/s_B strength pair. Population response patterns transitioned from being dominated by frontally tuned neurons when s_A was the stronger stimulus to being dominated by peripherally tuned neurons when s_B was the stronger stimulus, as expected (Fig. 2*B*, left to right). The responses of individual neurons were scaled and replotted to show more clearly the distribution of the shapes of response transitions with changing relative strengths (Fig. 2*C*).

If the OTid categorizes stimuli based on relative stimulus strength, we would expect that as relative stimulus strength gradually morphed, the pattern of population responses would not change gradually; rather, it would change abruptly from one pattern when $s_A > s_B$ to a different pattern when $s_A < s_B$. To test this expectation, we quantified the similarity between response patterns evoked by the different strength pairs (matrix columns) using the Pearson correlation coefficient (Fig. 2*D*). The analysis revealed that the evoked response patterns were indeed grouped: response patterns to strength pairs between 100%/0% and 56%/44% were similar, as were those between 44%/56% and 0%/100%, with an abrupt transition between 56%/44% and 44%/56% (Fig. 2*D*) (the corresponding speed pairs were 8.96°/s/7.04°/s and 7.04°/s/8.96°/s).

To evaluate whether the patterns of population responses were more similar within a category than between categories of relative stimulus strengths, we computed two parameters (Freedman and Assad, 2006): an average WCD and an average BCD in population response correlations for the 10 tested relative strengths of s_A and s_B (Materials and Methods). Equal values of WCD and BCD demonstrate that patterns of population responses are as similar to stimuli within categories as they are to stimuli between categories, thereby indicating a lack of categorization. On the other hand, a WCD value that is smaller than the BCD value demonstrates that patterns of population responses to stimuli within a category are more similar than those to stimuli between categories, thereby indicating categorization. As shown in Figure 2*E* (filled red circle), the WCD value (0.36 ± 0.04; mean ± SEM; Materials and Methods) calculated using the response correlation values in Figure 2*D* (bottom) was less than one-third of the BCD value (1.11 ± 0.04), demonstrating categorization. The strength of categorization, quantified as a scalar value using a categorization index, was 0.51; positive values of the index indicate smaller differences within than between categories (Materials and Methods). Thus, population responses in the OTid are largely insensitive to changes in relative stimulus strength as long as the same stimulus remains the strongest, and are particularly sensitive to changes in relative stimulus strength when a different stimulus becomes the strongest.

Population representation of the strength of a single stimulus is not categorical

Interestingly, although the population representation of the relative strength of multiple competing stimuli was categorical, the population representation of the strength of a single stimulus was not (Fig. 3). This was revealed by an analysis of the responses of frontally and peripherally tuned neurons to different strengths of s_A presented alone. For each frontally tuned neuron, s_A was presented in the center of its RF (Fig. 3*A*, top), while for each peripherally tuned neuron, s_A was presented outside the RF, 30° ipsilateral to the RF center (Fig. 3*A*, bottom). The strengths of s_A were chosen to be identical to those used in the strength morphing protocol. As before, a population response matrix was constructed by combining the responses of frontally ($n = 25$) and peripherally ($n = 20$) tuned neurons (Fig. 3*B*).

Correlation analysis demonstrated that, unlike in the strength morphing protocol, population responses to s_A alone changed gradually, rather than categorically, as the strength of s_A changed (Fig. 3*C*). Patterns of population responses to neighboring values of stimulus strength were similar, whereas those to well separated values of stimulus strength were poorly correlated (Fig. 3*C*, top) (high correlation values occurred only close to the diagonal). This is also clear from the horizontal transect through the correlation matrix at $s_A = 44\%$ (Fig. 3*C*, bottom). This lack of categorization by population responses to s_A alone was confirmed by the fact that the WCD and BCD values calculated from the response correlations (Fig. 3*C*, bottom) were nearly equal (Fig. 2*E*, blue open square).

Similarly, population responses to s_B alone did not exhibit categorization (Figs. 2*E*, green open square, and 3*D–F*).

In both the above cases, the responses of about half the neurons in the population were nearly zero for all strengths of the presented stimulus: all peripherally tuned neurons in Figure 3*B* and all frontally tuned neurons in Figure 3*E*. By contrast, in the strength-morphing protocol, the responses of most neurons changed as stimulus strengths varied: the responses of frontally tuned neurons typically decreased while those of the peripherally tuned neurons increased (Fig. 2*B*, left to right). Consequently, the changes in the responses of both subpopulations of neurons contributed to the large change in response correlation when stimulus strength pairs changed from $s_A > s_B$ to $s_A < s_B$ (Fig. 2*D*, bottom). We wondered whether this difference could account for the lack of categorization in the single stimulus protocol.

To test this possibility, we constructed a new population response matrix by combining the responses of frontally tuned neurons to s_A alone and the responses of peripherally tuned neurons to s_B alone, and organizing them such that the ordering of s_A and s_B strengths matched that in Figure 2*B* (Fig. 3*H*). In this matrix, the responses of frontally tuned neurons decreased from

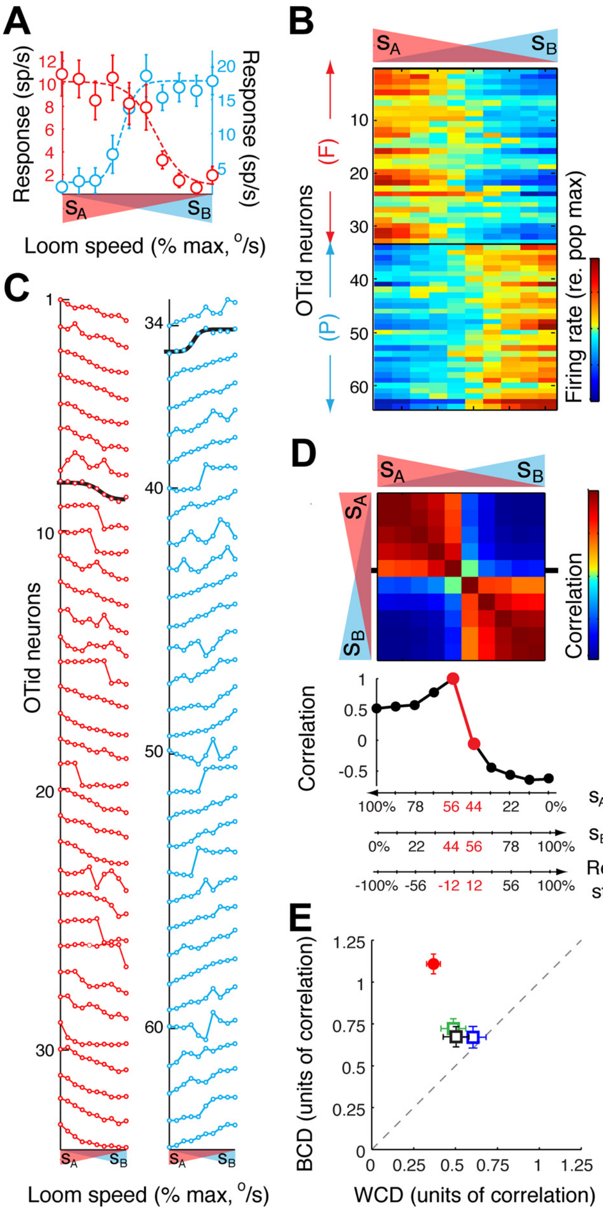

Figure 2. The population representation of relative stimulus strength is categorical. ***A***, Firing rate responses (mean ± SEM) to the strength-morphing protocol of one frontally tuned neuron (in red) and one peripherally tuned neuron (in blue). Dashed lines indicate best sigmoidal fits to data. Rates based on spikes counted over the 100–250 ms window after stimulus onset. Stimulus protocol is schematized below the *x*-axis with complementary gradients of loom speed ("strength") of s_A and s_B. Loom speeds are expressed as a percentage of maximum speed (16°/s; Materials and Methods). ***B***, Response patterns from 33 frontally tuned (F) and 31 peripherally tuned (P) neurons organized as a matrix (rows represent neurons, columns represent stimulus strength pairs). Firing rates of each neuron are mean-subtracted and normalized to the population maximum of the resulting responses. Therefore, positive values represent responses greater than the mean, while negative values represent responses less than the mean. ***C***, Responses of individual neurons (from ***B***) normalized to the range of responses of each neuron. Red, Frontally tuned neurons. Blue, Peripherally tuned neurons. Numbering along the *y*-axis is the same as in ***B***. Neurons #8 and 35 correspond to the two neurons shown in ***A***. Black curves are reproductions (from ***A***) of the best sigmoidal fits to the responses of these neurons. As relative stimulus strength changed, individual neuronal responses transitioned between response minimum and response maximum in different ways, from abrupt (e.g., neuron #10) to linear (e.g., neuron 36). For some neurons, responses did not change systematically as relative stimulus strength varied (e.g., neuron #50). ***D***, Top, Correlation matrix showing pairwise similarity between population response patterns evoked by the different stimuli (correlations between columns of matrix in ***B***). Bottom, Horizontal transect through the correlation matrix at the position indicated by tick marks. *x*-axis labels: s_A strength, s_B strength, and relative strength ($s_B - s_A$), expressed as percentage of maximum loom speed (Materials and Methods). Abrupt transition in population response patterns is seen when s_B is just equal to s_A: transition occurs between s_A/s_B strength pairs of 56%/44% and 44%/56%, i.e., between the speed pairs of 8.96°/s/7.04°/s and 7.04°/s/8.96°/s. ***E***, Quantification of the average difference in the response correlations for stimulus conditions within versus between categories. The categories are $s_A > s_B$ and $s_A < s_B$. WCD, Average within-category difference. BCD, Average between-category difference (Materials and Methods). The position of the filled red circle was calculated using correlation values shown in ***D***, bottom. The positions of the open squares were calculated using correlation values shown in the bottom panels of Figure 3, *C* (blue), *F* (green), and *I* (black).

←

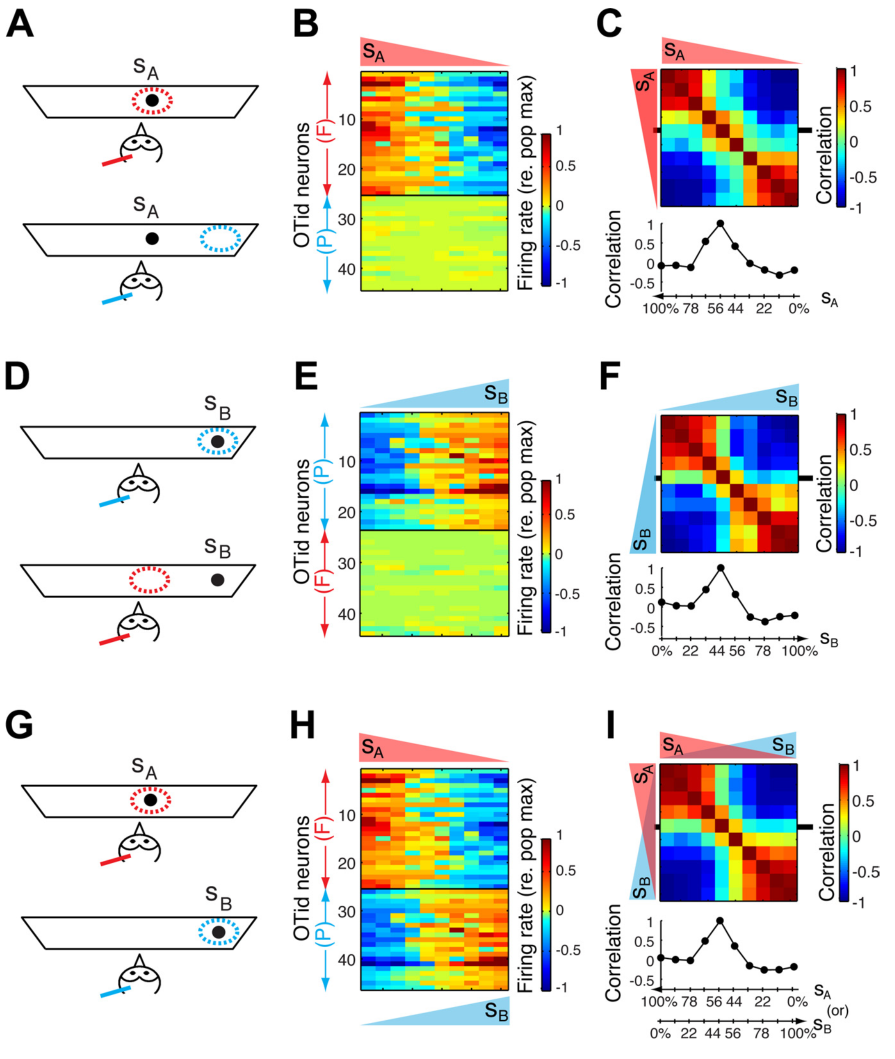

Figure 3. The population representation of the strength of a single stimulus is not categorical. Conventions in ***A***, ***D***, and ***G*** are as in Figure 1*B*; in ***B***, ***E***, and ***H***, they are as in Figure 2*B*; and in ***C***, ***F***, and ***I***, they are as in Figure 2*D*. ***A–C***, Responses to s_A alone. ***A***, Schematic of neuronal recordings from frontally (top) and peripherally tuned (bottom) neurons to s_A alone. ***B***, Population response patterns to s_A alone from 25 frontally tuned and 20 peripherally tuned neurons organized as a matrix (rows represent neurons, columns represent s_A strength). ***C***, Top, Correlation matrix showing pairwise similarity between population response patterns evoked by s_A (correlations between columns of matrix in ***B***). Bottom, Horizontal transect through the correlation matrix at the position indicated by tick marks. The pattern of responses changed gradually: only nearby s_A strengths exhibited similar values of correlation. ***D–F***, Responses to s_B alone (21 frontally tuned and 23 peripherally tuned neurons). ***F***, Population response patterns changed gradually. ***G–I***, Responses to s_A and s_B in the absence of competitive interactions did not show categorization. ***H***, We constructed a response matrix containing the responses of 25 frontally tuned neurons to s_A (same neurons/responses as in ***A–C***) and the responses of 23 peripherally tuned neurons to s_B (same neurons/responses as in ***D–F***). This matrix simulated population responses to the strength-morphing protocol in the absence of competitive interactions. ***I***, As in ***C*** and ***F***, patterns of population responses did not show abrupt transitions, indicating that the categorization observed in Figure 2 was the specific result of competitive interactions between the neural representations of s_A and s_B.

left to right (as the strength of s_A decreased), while those of peripherally tuned neurons increased from left to right (as the strength of s_B increased), similar to the matrix from the strength morphing protocol (Fig. 2*B*). Note that this response matrix represents the responses of the OTid population to the simultaneous presentation of s_A and s_B in the absence of competitive interactions. Correlation analysis showed that responses combined in this manner still did not result in a categorized representation of the strengths of single stimuli (Fig. 3*I*). Indeed, the WCD and BCD values computed from these response correlations (Fig. 3*I*, bottom) were still nearly equal in magnitude (Fig. 2*E*, black open square).

Thus, categorization by patterns of population responses in the OTid specifically reflects competitive interactions between the representations of multiple stimuli. Moreover, it indicates that the responses we describe here are not simply related to motor preparation for orienting, which would be categorical both in the single stimulus and the multiple stimulus cases.

A small subset of neurons mediates population-wide categorization

The abrupt transition in response pattern revealed in Figure 2*D* was abolished if 20% or more of the neurons with the most pronounced response transitions (highest absolute correlation with a template that represented an ideal categorizer) (Fig. 4*A*) were eliminated from the analysis (Fig. 4*B–D*). When only the top 10% of the neurons with the sharpest transitions were eliminated, the WCD was still much smaller than the BCD value (Fig. 4*E*, filled red circle). However, when the top 20% or 30% of these neurons were eliminated, the WCD and BCD values were nearly equal (Fig. 4*E*, open squares). Categorization by patterns of population responses in the OTid is therefore the result of coordinated changes in the responses of a small subset of neurons.

The category boundary is flexible

Finally, we investigated whether the category boundary observed in Figure 2*D* was flexible to changes in the strength of the strongest stimulus. We examined this by increasing the strength of s_A in all stimulus conditions by 22% (average = 3.5°/s ± 0.5°/s). In this protocol, the speed of s_A decreased from 19.5°/s to 3.5°/s, while the speed of s_B increased, as before, from 0°/s to 16°/s.

The responses of the example neurons in Figure 2*A* to this s_A-favored strength-morphing protocol are plotted in Figure 5*A*. Correlation analysis of population response patterns to the s_A-favored strength-morphing protocol revealed an abrupt transition (Fig. 5*B*), as before (Fig. 2*D*). However, the transition occurred at a new boundary, between the strength pairs of 66%/56% and 55%/67%, i.e., between the speed pairs of 10.54°/s/8.96°/s and 8.78°/s/10.72°/s. This new boundary was shifted to the right with respect to the old boundary, by an amount reflecting the systematic increase in s_A strength, and it occurred, as before, at the cross-over from $s_A > s_B$ to $s_A < s_B$. To estimate the strength of categorization, we calculated the WCD and BCD values using the response correlation values shown in Figure 5*B* (bottom) (see Materials and Methods). As in Figure 2*E*, the WCD value was much smaller than the BCD value (WCD = 0.38 ± 0.05; BCD = 1.06 ± 0.05), resulting in a positive categorization index of 0.47 (Fig. 5*C*, red). Thus, categorization by the OTid is flexible, the boundary shifting systematically with the strength of the strongest stimulus. Consequently, population responses in the OTid provide an explicitly

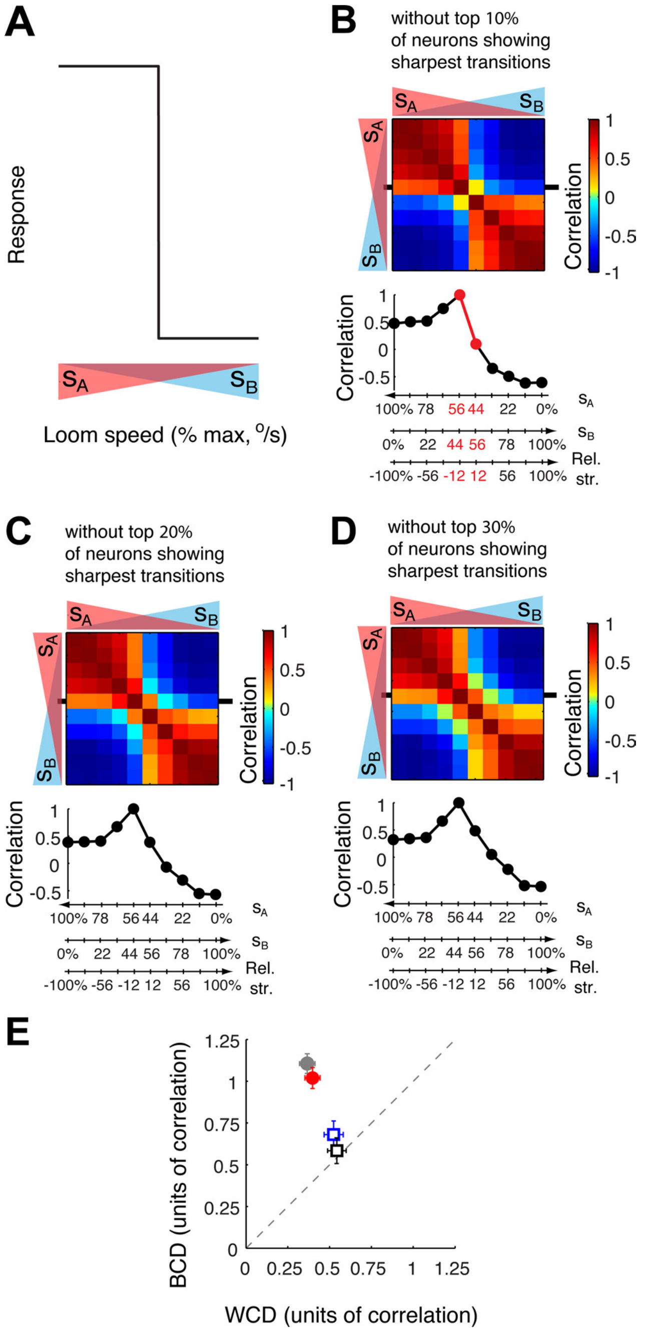

Figure 4. A small subset of neurons mediates population-wide categorization. ***A***, Template representing ideal categorizing responses of a neuron, with an abrupt transition occurring when $s_A = s_B$. ***B–D***, Conventions as in Figure 2*D*. Top, Correlation matrices (similarity between population response patterns) computed between the columns of the response matrix in Figure 2*B* after eliminating the top 10% (***B***), 20% (***C***), or 30% (***D***) of the neurons that exhibited the highest absolute correlation with the template in ***A***. Bottom, In ***B***, the abrupt transition between strength pairs 56%/44% and 44%/56% is still seen. However, in ***C*** and ***D***, abrupt transitions progressively diminish: nearby strength pairs always exhibit similar values of correlation (denoting gradually changing response patterns). ***E***, Quantification of the average difference in the response correlations for stimulus conditions within versus between categories (Materials and Methods). The categories are $s_A > s_B$ and $s_A < s_B$. The position of the filled red circle was calculated using correlation values shown in ***B***, bottom. The positions of the open squares were calculated using correlation values shown in the bottom panels of ***C*** (blue) and ***D*** (black). Filled gray circle: reproduction of the filled red circle in Figure 2*E*, for comparison.

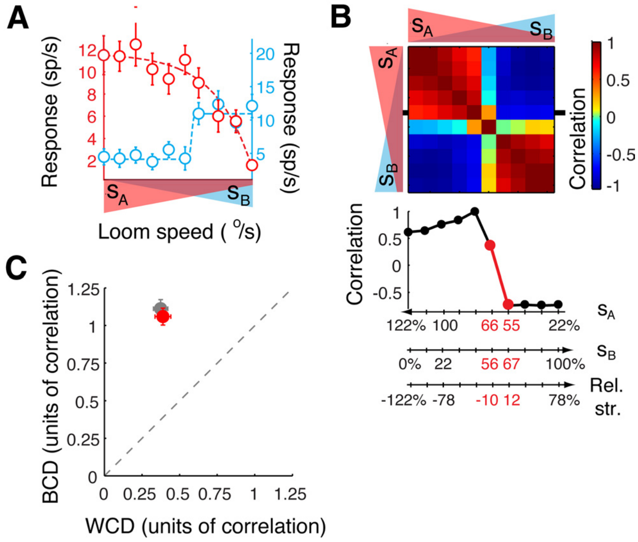

Figure 5. The category boundary for the population representation of the strongest stimulus is flexible. ***A***, ***B***, Responses to the s_A-favored strength-morphing protocol (Results). Conventions are as in Figure 2, *A* and *D*, respectively. ***A***, Firing rate responses of a frontally (red) and a peripherally (blue) tuned neuron (same as in Fig. 2*A*) show rightward shifts of response transitions. ***B***, Top, Correlation matrix showing pairwise similarity between population response patterns evoked by the different strength pairs (23 frontally and 26 peripherally tuned neurons). Bottom, Abrupt transition in the population response pattern occurred at a larger (right-shifted) s_B speed, corresponding, as in Figure 2*D*, to when s_B just exceeded s_A. The transition occurred between s_A/s_B strength pairs of 66%/56% and 55%/67%, i.e., between the speed pairs of 10.54°/s/8.96°/s and 8.78°/s/10.72°/s. ***C***, Quantification of the average difference in the response correlations for stimulus conditions within versus between categories (Materials and Methods). The categories are $s_A > s_B$ and $s_A < s_B$. The position of the filled red circle was calculated using correlation values shown in ***B***, bottom. Filled gray circle: reproduction of the filled red circle in Figure 2*E*, for comparison.

categorical representation of the location of the strongest stimulus.

Discussion

Bottom-up stimulus selection is integral to the decision of where to look or attend next, a brain function critical for survival. By reframing bottom-up stimulus selection as a categorization problem, we demonstrate that in the OTid, a structure in which this decision is known to be represented (Glimcher, 2001), categorical patterns of population responses explicitly and flexibly signal the strongest stimulus. The explicit nature of this representation (Gollisch and Meister, 2010) is evident in the observation that population response patterns within a relative strength category (for instance, $s_A > s_B$, or $s_B < s_A$) are similar, while they display abrupt transitions between categories (Fig. 2). As a result, the pattern of inputs to neurons downstream from this population is similar across all stimulus conditions within a category, and markedly different between categories, thereby enabling category-specific responses. The flexibility of the category boundary is evident in the observation that modifying the strength of one of the stimuli so that strength equality occurs at different absolute values of stimulus strengths results in a corresponding shift in the categorization boundary (Fig. 5). Thus, the OTid explicitly and flexibly categorizes input based on relative stimulus strength.

Flexible categorization and "switch-like" responses

In previous work exploring the representation of relative stimulus strength in the OTid (Mysore et al., 2011), we showed that individual neurons exhibit a wide range of sensitivities to changing relative stimulus strengths. Relative stimulus strength in that study was varied by fixing the strength of a stimulus inside the RF, while varying that of a second, competing stimulus outside the RF. We showed that a subset of neurons, representing about 30% of the tested population, exhibited sharp, "switch-like" decrements in response as the strength of the stimulus outside the RF exceeded that of the stimulus inside the RF, while the remaining neurons exhibited more gradually changing responses or responses uncorrelated with relative stimulus strength. A key question not addressed therein was how the responses of these switch-like neurons might be read out by a downstream decoder to achieve reliable identification of the strongest stimulus. One possibility was that decoder neurons had access selectively to OTid neurons with switch-like responses. Though the efficacy of such a decoder has been demonstrated (Mysore et al., 2011), it is unclear whether such decoder neurons actually exist.

Here, by analyzing the patterns of population responses in the OTid constructed from all tested neurons regardless of the nature of their response transitions, we have demonstrated that no such selective access is necessary: the population responses of the OTid, "as a whole," are able to categorically signal the strongest stimulus. The ability of the OTid to categorize relative stimulus strength depends on 20% of the neurons that exhibit the most pronounced transitions, i.e., neurons well within the "switch-like" population. This indicates that while the decoder must receive switch-like input to achieve explicit categorization, it need not receive only switch-like input to do so. Such population-wide categorization that depends on the response properties of a small subset of neurons has been observed in other systems as well (Meyers et al., 2008; Niessing and Friedrich, 2010).

The only distinctive functional property of the subset of neurons critically involved in explicit categorization was the abrupt, switch-like transitions in their responses to competing stimuli of varying relative stimulus strengths. In a previous study, we compared RF size, RF location, responses to increasing strengths of single stimuli, and spike widths (Mysore et al., 2011) between neurons with switch-like responses and those with non-switch-like responses and found no difference in these properties between the two subsets. This suggests that the essential property that distinguishes switch-like neurons from all others is the projections they receive from inhibitory elements that mediate switch-like suppression.

Finally, the results presented here establish the response properties that are necessary and sufficient for explicit, flexible categorization of relative stimulus strength. First, a subset of neurons must display switch-like responses as the relative strength of the competing stimuli changes. Second, the value at which the switch-like decrease in responses occurs must shift systematically with changes in the strength of the strongest stimulus.

Categorization and the properties of competing stimuli

Categorization should occur across all locations and all stimuli that are relevant to the species. Here, we tested categorization with one stimulus located frontally and another located peripherally. For stimuli that are located within the same hemifield, the strength of global suppression does not change significantly as long as the competing stimulus is located outside the RF (Mysore et al., 2010). This suggests that switch-like responses, and therefore explicit categorization, will occur for all competing stimuli within the same hemifield. For stimuli that are located in opposite hemifields, response suppression still exists, but the magnitude of the suppression is substantially less [approximately one-third the strength (Mysore et al., 2010)]. Assuming that the mechanisms underlying switch-like responses also operate across hemifields, we would expect to observe explicit categorization in this case as well, with the strength of the categorical responses reduced correspondingly.

With respect to the sensory modalities of the competing stimuli, our experiments involved the use of only visual stimuli. However, switch-like responses occur even when the competing stimulus is an auditory noise burst (Mysore et al., 2011). These results indicate that a categorical representation of the strongest stimulus will occur independently of the nature of the competing stimuli.

The range of stimulus strengths within which categorization should occur must correspond to the range of strengths relevant to the species. The range of loom speeds tested in this study (0°/s to 21°/s) corresponds to the range of loom speeds that most neurons in the OTid encode with increasing firing rates (Mysore et al., 2010, 2011), suggesting that this range of speeds is behaviorally relevant to barn owls. We demonstrated categorization at loom speeds of 8°/s and 9.75°/s. We predict that categorization will also occur at all other loom speeds within the range. Together, these arguments suggest that categorization would occur independently of the spatial locations of the stimuli, independently of the sensory modalities of the stimuli, and over ranges of stimulus strengths behaviorally relevant to the species.

Categorization of relative strength: not just motor responses

The SCid/OTid has a well established role in encoding motor plans and driving orienting behavior (Wurtz and Albano, 1980). Is it possible that the categorization reported here is the trivial result of measuring motor-related activity in the OTid? This possibility can be ruled out because the categorization observed resulted specifically from competition between representations of multiple stimuli (Fig. 2 vs Fig. 3); the responses to single stimuli were not categorical. Had the responses been the activity of saccade-related neurons, the responses to single stimuli should have been categorical as well. Moreover, switch-like responses are seen even in tranquilized, untrained animals (Mysore et al., 2011). Thus, while the categorization in the OTid most likely plays a fundamental role in competitive selection, it represents a step before the commitment to a choice that results in directing attention and/or the execution of a motor plan.

Categorization and behavior

The categorization of relative stimulus strength observed here in the owl OTid could explain the stimulus selection deficits that have been observed following the inactivation of the SCid in mammals. Recent studies in primates and rodents have shown that selection of a target stimulus among competing distracter stimuli is impaired following SCid inactivation, but that the selection of a single stimulus is not (McPeek and Keller, 2004; Felsen and Mainen, 2008; Lovejoy and Krauzlis, 2010; Nummela and Krauzlis, 2010). Importantly, the impairment with multiple stimuli increases dramatically as the difficulty of discriminating among the stimuli increases. A fundamental computational advantage of categorization is that it produces large separations between output patterns even for inputs from different categories that are very close together (for instance, $s_B = 44\%$ vs $s_B = 56\%$, in Fig. 2*D*; and $s_B = 56\%$ vs $s_B = 67\%$, in Fig. 5*B*). Consequently, such categorized representations, as demonstrated here for the owl, could account for the improved performance in intact versus SCid-inactivated animals performing difficult selection tasks (Mysore and Knudsen, 2011).

Explicit versus implicit categorization

Whereas the OTid explicitly signals the strongest stimulus, other brain areas implicated in this decision process, such as the LIP and FEF, have, thus far, been shown to represent relative strength-dependent categories implicitly. In these other structures, neural responses to different relative stimulus strengths change gradually, rather than categorically, but they can be used to create categories with further downstream computations (Bisley and Goldberg, 2003; Ferrera et al., 2009). An experimental protocol similar to the one used here may well reveal explicit strength-based categorization in those brain areas as well. If, however, those brain areas were to be shown to encode relative strength-based categories only implicitly, it would indicate that the OTid (SCid) is capable of additional computations in the service of bottom-up stimulus selection. A similar conclusion has been drawn for the representation of categories of visual motion direction in the LIP (explicitly categorical) versus in the middle temporal area [not explicitly categorical (Freedman and Assad, 2006)].

Flexible categorization versus winner-take-all

Stimulus selection for attention and gaze is generally thought to operate via a winner-take-all mechanism (Koch and Ullman, 1985). However, representations of relative stimulus strength in the OTid do not exhibit the signature property of winner-take-all computations, i.e., responses to the "losing" stimulus being driven to zero (Mysore et al., 2011). Instead, OTid responses to the losing stimulus scale with the absolute strength of the losing stimulus. Nonetheless, explicit and flexible categorization in the OTid shares the other three properties of a classic winner-take-all computation: responses within a category are similar; responses between categories are markedly different; and the category boundary shifts with the strength of the winning stimulus. Explicit flexible categorization is a more general computation that subsumes the special case of the winner-take-all computation. Thus, although not winner-take-all, explicit flexible categorization nonetheless successfully achieves the signaling of the strongest stimulus.

Categorization for "what" versus categorization for "where"

Previous examples of explicit and flexible categorization reported in several brain areas (zebrafish olfactory bulb, songbird high vocal center, and monkey prefrontal cortex and LIP) all show that neurons classify stimuli to signal "what" the stimulus is (e.g., odor A vs odor B, dog vs cat, etc.) (Freedman et al., 2001; Freedman and Assad, 2006; Prather et al., 2009; Niessing and Friedrich, 2010). In contrast, categorization in the OTid signals "where" the strongest stimulus is. Remarkably, the strength of categorization, as quantified by using the categorization index, is higher in the OTid of passive, untrained owls (0.51) than those reported in either the prefrontal (mean = 0.2) or the parietal (mean = 0.27) cortices in monkeys trained extensively to perform categorization of stimulus identity (Freedman et al., 2001, 2006). Moreover, whereas categorization described in most of those brain areas is learned, categorization by the OTid [as in the olfactory bulb (Niessing and Friedrich, 2010)] is innate, likely the result of evolutionary tuning expressed in the wiring of the midbrain selection network (Knudsen, 2011). Our findings suggest the intriguing possibility that stimulus selection for attention and gaze shares neural mechanisms with categorization for identification, and perhaps with decision making (Freedman and Assad, 2011).

References

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

Carello CD, Krauzlis RJ (2004) Manipulating intent: evidence for a causal role of the superior colliculus in target selection. Neuron 43:575–583.

Fee MS, Mitra PP, Kleinfeld D (1996) Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. J Neurosci Methods 69:175–188.

Felsen G, Mainen ZF (2008) Neural substrates of sensory-guided locomotor decisions in the rat superior colliculus. Neuron 60:137–148.

Ferrera VP, Yanike M, Cassanello C (2009) Frontal eye field neurons signal changes in decision criteria. Nat Neurosci 12:1458–1462.

Freedman DJ, Assad JA (2006) Experience-dependent representation of visual categories in parietal cortex. Nature 443:85–88.

Freedman DJ, Assad JA (2011) A proposed common neural mechanism for categorization and perceptual decisions. Nat Neurosci 14:143–146.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291:312–316.

Glimcher PW (2001) Making choices: the neurophysiology of visual-saccadic decision making. Trends Neurosci 24:654–659.

Gollisch T, Meister M (2010) Eye smarter than scientists believed: neural computations in circuits of the retina. Neuron 65:150–164.

Itti L, Koch C (2001) Computational modelling of visual attention. Nat Rev Neurosci 2:194–203.

Knudsen E (2011) Control from below: the contribution of a midbrain network to spatial attention. Eur J Neurosci, in press.

Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4:219–227.

Leopold DA, Logothetis NK (1999) Multistable phenomena: changing views in perception. Trends Cogn Sci 3:254–264.

Lovejoy LP, Krauzlis RJ (2010) Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. Nat Neurosci 13:261–266.

McPeek RM, Keller EL (2004) Deficits in saccade target selection after inactivation of superior colliculus. Nat Neurosci 7:757–763.

Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T (2008) Dynamic population coding of category information in inferior temporal and prefrontal cortex. J Neurophysiol 100:1407–1419.

Mitra P, Bokil H (2008) Observed brain dynamics. New York: Oxford UP.

Müller JR, Philiastides MG, Newsome WT (2005) Microstimulation of the superior colliculus focuses attention without moving the eyes. Proc Natl Acad Sci U S A 102:524–529.

Mysore SP, Knudsen EI (2011) The role of a midbrain network in competitive stimulus selection. Curr Opin Neurobiol, in press.

Mysore SP, Asadollahi A, Knudsen EI (2010) Global inhibition and stimulus competition in the owl optic tectum. J Neurosci 30:1727–1738.

Mysore SP, Asadollahi A, Knudsen EI (2011) Signaling of the strongest stimulus in the owl optic tectum. J Neurosci 31:5186–5196.

Niessing J, Friedrich RW (2010) Olfactory pattern classification by discrete neuronal network states. Nature 465:47–52.

Nummela SU, Krauzlis RJ (2010) Inactivation of primate superior colliculus biases target choice for smooth pursuit, saccades, and button press responses. J Neurophysiol 104:1538–1548.

Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R (2009) Neural correlates of categorical perception in learned vocal communication. Nat Neurosci 12:221–228.

Wang XJ (2008) Decision making in recurrent neuronal circuits. Neuron 60:215–234.

Wurtz RH, Albano JE (1980) Visual-motor function of the primate superior colliculus. Annu Rev Neurosci 3:189–226.