Neural circuit mechanism for generating categorical representations

Authors: Nagaraj R. Mahajan[1] and Shreesh P. Mysore[2,3]*

[1]Department of Electrical and Computer Engineering, Johns Hopkins University

[2]Departments of Psychological and Brain Sciences, Johns Hopkins University

[3]The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University

*Correspondence to: shreesh.mysore@jhu.edu

Abstract: Categorical neural representations underlie various forms of selection and decision-making. They promote robust signaling of the winner in the presence of input ambiguity and neural noise. Here, we show that a 'donut-like' inhibitory mechanism in which each competing option suppresses all options except itself is highly effective at generating categorical responses. It surpasses motifs of feedback inhibition, recurrent excitation, and divisive normalization used in decision-making models. We demonstrate experimentally not only that this mechanism operates in the midbrain spatial selection network in barn owls, but also that it is required for categorical signaling by it. Moreover, the pattern of inhibition forms an exquisitely structured 'multi-holed' donut consistent with this network's combinatorial inhibitory function. Self-sparing inhibition may be a powerful circuit module central to categorical selection.

Main Text: Selection among competing options is a fundamental component of adaptive behavior. Categorical neural representations have been reported in multiple brain areas to underlie selection and decision-making across animal species (*1-10*). These representations, which involve the transformation of continuously varying inputs into discrete output classes, involve a large, abrupt change in responsiveness across the selection boundary (Fig. 1A – red). They enhance the reliability of neural signaling of the selected category (or 'winner') specifically when competing options on opposite sides of the selection boundary are very similar to one another (i.e., when there is input ambiguity), and in the face of neural response variability (i.e., neural 'noise') (Fig. 1A-red vs. orange). This robustness to sensory and neural noise of the decoding of the winner from neural responses can be quantified using boundary discriminability (*bd'*) - the discriminability (d-prime; Methods) between two competing options on either side of the selection boundary: bd' is higher for more categorical response profiles (Fig. 1A and 1B-left, open symbols: red vs. orange). However, *bd'* is higher also for response profiles that are more robust-to-noise *without* being more categorical - for instance, scaled response profiles (Fig. 1A-orange vs blue). A commonly used metric that is insensitive to scaling and that captures, specifically, the strength of categorization by response profiles is the categorization index ((*2, 5*); it operates effectively on the normalized response profiles; Fig. 1B-middle inset). Here, we define a modified version, *CatI*, that takes into account response variability as well: it compares the average between-category discriminability with the average within-category discriminability (Methods; Fig. 1B-right, filled symbols). Thus, robustness to noise (*bd'*) can be improved both by scaling up responses and by making them more categorical (higher *CatI*; Fig. 1AB, black vs. red) However, the limited dynamic ranges of neural responses (which limit the range of scaling possible), as well as the inherent discreteness of categorical neural representations, highlights

their importance for promoting robust-to-noise selection. Despite the clear utility of categorical representations, their pervasiveness in brain areas (*1-10*), as well as the range of theoretical formulations and computational models proposed to account for them (*11-14*), neural circuit mechanisms that control their production have remained elusive. Specifically, it is unclear what aspects of circuit architecture are essential for generating categorical responses.

Donut-like inhibitory motif controls categorical signaling in model circuits

We reasoned that properties of neural inhibition, which is known to be involved in comparing representations of competing options (*15-18*), would impact the nature of competitive response profiles. This led to two concrete hypotheses re circuit mechanisms. The first was that feedback inhibition between the options, i.e., inhibition whose strength is iteratively influenced by both options, may play a key role. This was motivated by extensive modeling of categorical decision making (*13, 19*), work on direction selectivity in the retina (*20*), as well as work on spatial selection in barn owls (*21*). The second hypothesis was that a donut-like pattern of competitive inhibition, i.e., one in which each option suppresses others more strongly than it suppresses itself, may play a key role. This was motivated by work in turtles (*22*) as well as by the need, in categorical profiles, for enhanced response differences across the selection boundary: we reasoned that having strong inhibition to 'other' options, but weak "self" inhibition may enhance response differences.

To assess the computational validity as well as relative efficacies of these hypothesized circuit mechanisms, we turned to modeling. Starting with a generic baseline circuit capable of comparing the representations of competing options (Fig. 1C- model on top-left; Methods;(*16-18*)), we examined the impact of introducing just feedback inhibition (Fig. 1C- bottom-left), just donut-like inhibition (Fig. 1C- top-right), or both (Fig. 1C- bottom-right), on the production of

categorical response profiles in the presence of noise. Response profiles were measured to a classic two-stimulus morphing protocol in which the relative strength of the two stimuli (S1 and S2) was systematically varied, resulting in two categories (S1> S2 and S1<S2; Fig. 1E-bottom, graphic; (*2*, *4*, *5*)). Model neurons in the circuit were simulated with noisy, sigmoidal input-output functions (fano factor of 6; 30 repetitions per 'neuron', n=50 'neurons'; Methods). The resulting response profiles from the four models clearly involved a scaling component (Fig. 1E-left**)**, and therefore, we compared both scaling-dependent robustness-to-noise (*bd'*) as well as the scaling-insensitive categorization strength (*CatI)* across the response profiles (ANOVA with HBMC correction; Methods).

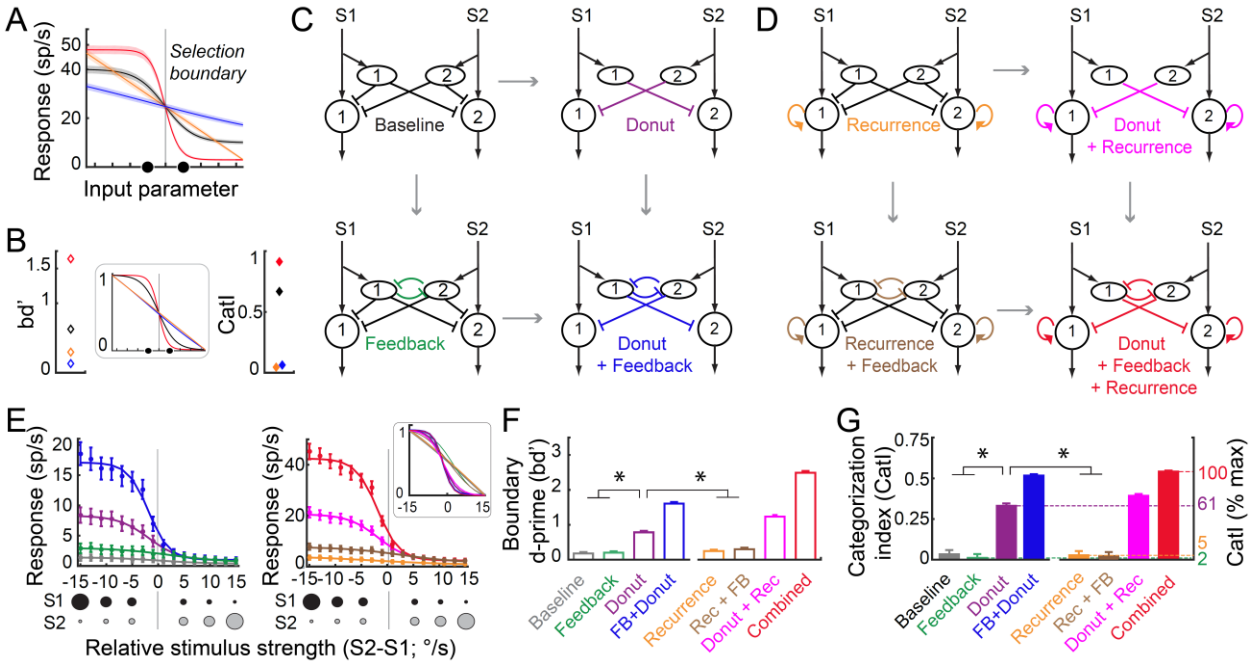


Fig. 1. Modeling: Donut-like inhibition surpasses other hypothesized circuit motifs in generating categorical responses to competing stimuli.

(A) *Schematic showing 4 different mathematically generated response profiles, as a function of continuously varying input. Vertical line: categorization or selection boundary. Translucent band: variability in responses; fano factor of 6 used to generate these responses (Methods). Black filled dots: Two competing inputs just straddling the selection boundary: nearly equal in input value (at a distance of 3 units from the selection boundary), but belonging to different categories.*

(B) *Left (open symbols): Boundary discriminability (bd'); defined as d-prime between the responses to the two competing inputs indicated in A (Methods), computed for each response profile in A (colors). Middle-inset: Response profiles in A normalized between 0 and 1; only means are shown; red profile transitions more abruptly than black, orange and blue profiles are equally gradual (linear). Right (filled symbols): Categorization index (CatI; Methods) for response profiles in A. (2, 5).*

(C, D) *Computational models incorporating different circuit motifs (Methods; Fig. S1). All models built upon generic baseline circuit capable of comparing competing options (C; top-left). Each model shows two 'channels'; each channel is the group of neurons (numbered) involved in representing a stimulus (S1 or S2); the goal of these models is to signal if S1>S2 (category 'A') or S2>S1 (category 'B'). Circles - output neurons; ovals - inhibitory neurons; arrows with pointed heads - excitatory connections, arrows with flat heads - inhibitory connections. (C) Four circuit models: baseline circuit (top-left), inclusion of feedback inhibition (bottom-left), donut-like inhibition (top-right) or both (bottom-right; Methods). (D) Four more circuit models: baseline circuit with recurrent amplification in each channel (top-left), followed by inclusion of feedback inhibition (bottom-left), or donut-like inhibition (top-right) or both (bottom-right).*

(E) *Simulated response profiles of 'example' output neuron 1 from each of the models in C (left) and D (right); obtained using a two-stimulus (S1 and S2) morphing protocol (bottom inset; Methods). Responses are mean ± s.e.m of 30 repetitions. The continuously varied input parameter was the relative strength of the two stimuli (S2-S1). Lines – best sigmoidal fits. Bottom schematic: Morphing protocol; stimulus strength schematized by size of dot. Input-output functions of model neurons were sigmoids with Gaussian noise (Methods; fano factor=6). Right-Inset: Response profiles normalized between 0 and 1 (same conventions as B-inset).*

(F, G) *'Population' summary of bd' (F) and CatI (G) of response profiles from various circuit models (colors); mean ± s.e.m.; n=50 model neurons. '*': p<0.05, ANOVA followed by Holm-Bonferroni correction for multiple paired comparisons; only a key subset of significant differences indicated for clarity.*

See also Fig. S1.

We found that feedback inhibition between competing channels had no effect on *bd'* (Fig. 1EF – green vs. grey; p= 0.99) or on *CatI* (Fig. 1E-inset and G – green vs. grey; p =0.99), and this result was largely independent of the strength of feedback inhibition (Fig. S1A). By contrast, introduction of a donut-like pattern of inhibition provided a boost in robustness-to-noise (Fig. 1EF – purple vs. grey; p =5.99 e-8), and an even greater boost in the strength of categorization of the response profiles (Fig. 1E-inset and G – purple vs. grey; p = 5.98e-8). The magnitudes of improvement were inversely related to the strength of "self" inhibition, reaching the maximum when "self" inhibition was zero (Fig. S1B; *bd'*: $\rho = -0.89$, $p = 1.64\ e\text{-}4$; *CatI*: $\rho = -0.81$, $p = 2.6\ e\text{-}3$ Pearson correlation test), thereby revealing the importance of donut-like inhibition over feedback inhibition for categorical responses. These results held true also when we started with a

circuit that included recurrent excitation, a common element in models of decision-making thought to play a role in categorical selection (Fig. 1D, and E-G: right panels, Fig. S1C; (*13, 23, 24*)).

Donut-like inhibition, therefore, emerged as the most important single motif for producing categorical responses (Fig. 1G, right axis; 61% of combined *CatI*), and this held true over the range of fano factor values tested (Fig. S1D). Whereas the presence of feedback inhibition or both recurrent excitation and feedback inhibition enhanced the impact of the donut-like motif (Fig. 1FG: blue vs. purple, red vs. purple; $p < 6.023$ e-8 in both cases), without the donut-like motif, they were nearly ineffective, either individually or together, at signaling the winning option categorically in the presence of input ambiguity and neural noise (Fig. 1FG: green, orange, brown).

Donut-like inhibitory motif in barn owl midbrain

We wondered, next, if the donut-like inhibitory motif operates within a neural selection circuit. An excellent site in the brain to investigate this is the midbrain spatial selection network in vertebrates (*25, 26*). It includes the superior colliculus, SC (optic tectum, OT, in non-mammals), a sensorimotor hub that is required for target selection for spatial attention (*27, 28*). Neurons in the intermediate and deep layers of the SC (SCid/OTid) encode a topographic spatial map of stimulus priority (priority = physical salience + behavioral relevance) (*29*), and OTid neurons signal the highest priority stimulus categorically*:* they respond with a high firing rate when the stimulus inside their RF is the one with highest priority, but their responses are suppressed to a low level when the RF stimulus is no longer the one with highest priority (*5, 6, 30*). The source of this competitive response suppression in the OTid is a group of GABAergic midbrain tegmental neurons called Imc in birds (isthmi pars magnocellularis; (*31*)): Imc inactivation

abolishes all competitive interactions in the OTid and thereby its ability to signal the winner (*32*, *33*).

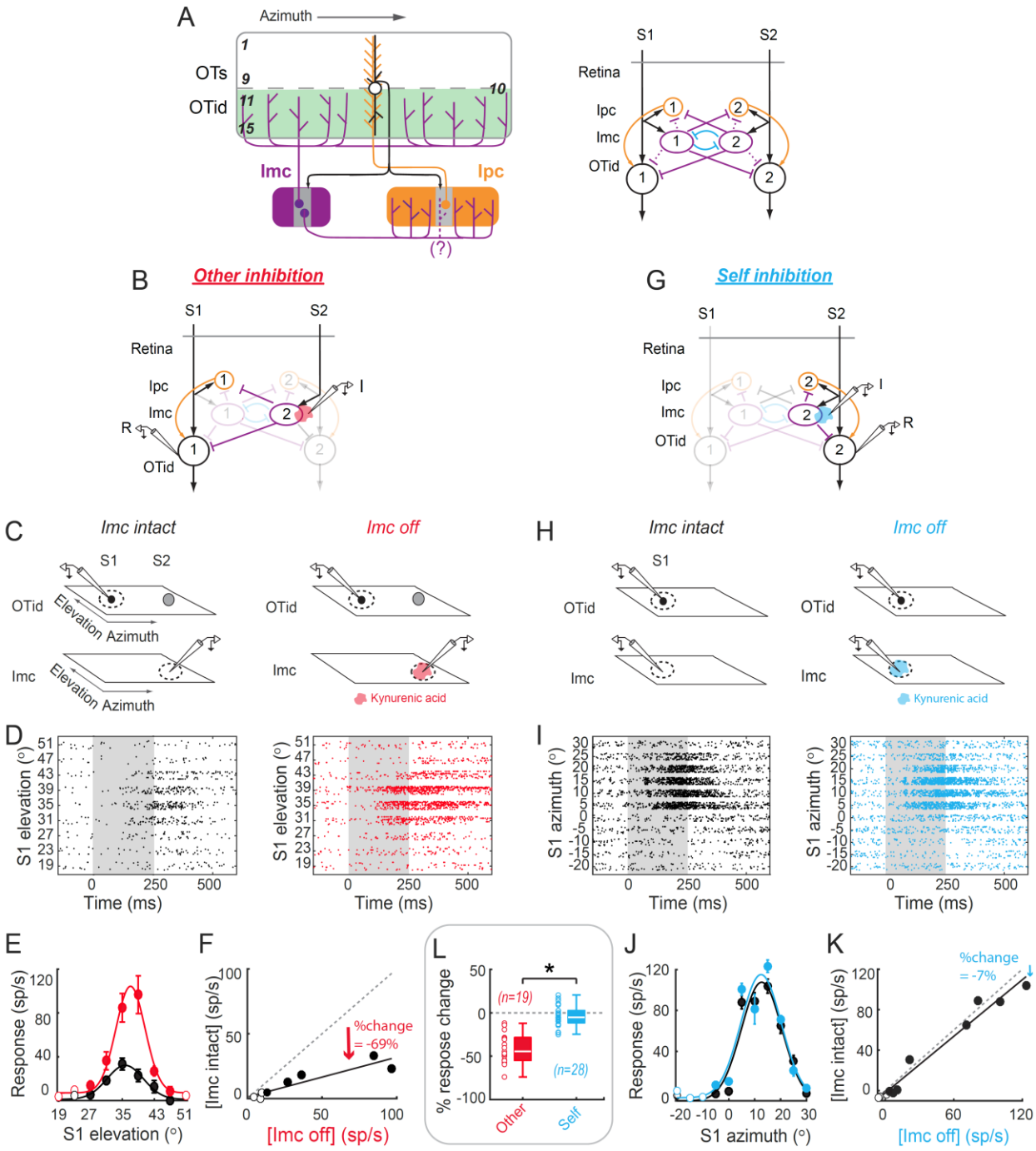


Fig. 2. Barn owl midbrain selection network contains functional donut-like inhibitory motif operating along azimuthal space.

(A) *Schematic of avian midbrain selection network. Left: Superficial (OTs), and intermediate-deep (OTid) layers of optic tectum; numbers: individual layers (1-15). Columns across layers of OT tissue, from left to right, encode individual locations in space (here, azimuth) topographically (34). Imc (purple): Isthmi pars magnocellularis, inhibitory neurons in midbrain tegmentum. Ipc (orange): Isthmi pars parvocellularis, cholinergic neurons. Imc and Ipc receive focal input from OT10 (dark grey projections). Ipc projects back focally to OT (orange projections). Imc projects back broadly across OTid and Ipc (purple projections); thought to spare the portion of OT providing input (absence of purple projections in the column occupied by black OT10 neuron; '?' (31): unknown if congruent portion of Ipc space map is also spared of Imc projections. Right: Network model showing the OT-Imc-Ipc circuit; conventions as in Fig. 1. Simulations from this model in Fig. S2. Dashed lines: Unknown if these connections exist functionally.*

(B-F) *Measurement of the strength of net "other" inhibition from Imc → OTid with paired recordings in Imc and OTid. "Net" indicates the combined inhibition due to both the direct (Imc → OTid) and indirect (Imc → Ipc → OTid) pathways (text). "Other" indicates that the OTid neuron encodes for (distant) spatial locations outside Imc neuron's RF. (B) Experimental setup. Iontophoresis and recording electrode (I) in the portion of Imc (encoding for stimulus S2); recording electrode (R) in the portion of OTid encoding for distant location (and stimulus S1). (C) Schematic of OTid and Imc space maps (quadrilaterals) showing RFs of neurons being recorded (dotted ovals) and stimulus protocol (black filled dot – S1; gray filled dot – S2). S1 and S2 are looming visual stimuli of fixed contrast but different loom speeds (strength; Mysore*

2010); S1 = 9.6 °/s, S2 = 19.2°/s (Methods). (D) Raster responses of example OTid neuron to paired stimulus protocol in C, in the Imc-intact condition (left column; black data) and Imc-off condition (right column – red data). Imc inactivation by (reversible) iontophoresis of kynurenic acid (a pan-glutamate receptor blocker;(32); Methods). Gray shading: stimulus duration (E) Response firing rates of this OTid (computed from rasters in D; spike count window = 100-350 ms); mean ± s.e.m. Lines: best Gaussian fits. Filled dots: responses to S1 at locations inside the OTid RF; open circles, outside RF (Methods). (F) Scatter plot showing OTid responses in Imc-intact vs. Imc-off conditions. Line: Best fit straight line; slope = 0.31, r^2 =0.76. % change in responses directly estimates the strength of suppression at this OTid neuron due to Imc (here, 69%; 100(1-0.31)).*

(G-K) *Measurement of the strength of net "self" inhibition from Imc → OTid with paired recordings in Imc and OTid. (conventions as in B-F). (G, H) OTid neuron is spatially aligned with Imc neuron (both encode overlapping locations); distance between OTid and Imc RF centers = 1.5°. (K) Line: Best fit straight line; slope = 0.93, R^2 =0.95. % change in responses directly estimates the strength of suppression at this OTid neuron due to Imc (here, 7%).* ***(L)*** *Population summary of strength of net "other" inhibition (red; n=19 Imc-OTid pairs), and strength of net "self" inhibition (blue; n=28 pairs) from Imc → OTid. Average distance between OTid and Imc RF centers in "other" experiments = 26.8° +/- 2.3°; in "self" experiments = 2.9° +/- 0.7 °. p<1e-5 (red vs. blue), p=<1e-5 (red vs 0), p=0.12 (blue vs. 0), paired t-tests with HBMC correction (Methods).*

See also Fig. S2.

This network has been shown to contain both feedback inhibition among competing channels and recurrent excitation within each channel. It implements the former as direct, long-range inhibition between Imc neurons (Fig. 2A-right – blue; not shown in 2A-left for clarity; (*31*, *35*)), and the latter, as response amplification by a group of cholinergic neurons, Ipc (Fig. 2A; orange), that connect in a point-to-point manner with the OT (*36*-*38*).

What, however, of the donut-like motif? Imc neurons are known to receive input from a focal portion of the OT (within layer 10; OT_{10}), but to suppress OTid activity broadly across the space map by sending projections via two pathways (Fig. 2A): (a) directly to the OTid (layers 11-15) (Fig. 2A – purple;(*31*)), and (b) indirectly to the OTid by inhibiting the potent point-to-point amplifiers of OTid activity, namely the cholinergic Ipc (Fig. 2A –orange; (*31*, *33*)). Incidentally, simulations with a circuit model in which the recurrent amplifier is also the recipient of powerful inhibition, as in the OT-Imc-Ipc circuit, confirm the superiority of the donut-like motif for producing categorical representations (Fig. S2A-C).

Whereas anatomical tracing studies (*31*) have indicated that the direct projections from the Imc to the OT may spare the portion of the OT providing input to the Imc (Fig. 2A; absence of purple projections in portion of OT around black neuron), this not been established functionally. Crucially, whether or not the indirect inhibitory pathway involving the Ipc exhibits the donut-like motif is unknown (Fig. 2A, purple '?'). This is critical because this indirect pathway is known to constitute the dominant route of inhibition from Imc to OTid (*31*): the majority of Imc projections target the Ipc (rather than the OTid; (*31*)), and Ipc provides substantial amplification of OTid activity (1.43x, on average; (*38*)).

To address functionally whether the Imc-Ipc-OT circuit implements a donut-like pattern of competitive inhibition, we made dual extracellular recordings in the barn owl OTid and Imc

(Methods; Fig. 2BG). We did so to measure directly the strength of the net competitive inhibition delivered by Imc neurons onto OTid neurons encoding for stimuli at distant locations ("other" inhibition), and separately, the strength of net inhibition delivered by Imc onto OTid neurons encoding for overlapping locations ("self" inhibition), (Fig. 2BH).

To measure the strength of "other" inhibition, we recorded the responses of OTid neurons to a stimulus (S1) inside their receptive field (RF; Fig. 2B, C-left; Methods) while simultaneously presenting a competing stimulus outside the RF (S2; at a distant azimuthal location from S1; Methods). Responses to S1 in the OTid are known to be suppressed by a distant S2, by an amount depending on their relative strength (*15, 39*). Notably, this suppression is known to be abolished upon focally inactivating the portion of Imc representing S2, revealing Imc as the primary source of competitive suppression in the OTid (*32, 33*). Therefore, to estimate the strength of inhibition delivered by Imc neurons to OTid neurons encoding 'other' locations, we compared the responses to paired presentation of S1 and S2 when Imc was intact (Fig. 2C-left) versus when the portion of Imc encoding S2 (spatially mismatched with the OTid recording site), was focally and reversibly inactivated (Fig. 2C-right). Inactivation was achieved by the blockade of excitatory synaptic inputs onto Imc neurons through iontophoresis of kynurenic acid (Methods). The difference in OTid responses between the Imc-intact (Fig. 2D-left; 2E-black) and the Imc-inactivated (Fig. 2D-right; 2E-red) conditions quantified the amount of suppression due to Imc onto the "other" OTid location, thereby directly estimating the strength of net "other" inhibition. (No change in OTid responses would indicate zero competitive suppression by Imc onto that OTid neuron.)

We found that Imc neurons exerted strong inhibition onto OTid neurons encoding for distant, non-matched spatial locations (Fig. 2F, L-red; mean strength = - 40.47 % +/- 17.70 %,

n==19 paired neurons; p=9.43e-9, t-test against 0; mean distance between centers = 26.74 °). We verified that the results were specifically due to Imc inactivation by observing that OTid responses to paired S1 and S2 after recovery from iontophoresis (measured 15 min after the drug was turned off; Methods) returned to pre-drug levels (Fig. S2D). Indeed, the suppression provided by Imc accounted for nearly all the suppression exerted by the stimulus S2 (Fig. S2E). In these experiments, Imc was inactivated effectively (median = 95%, 95% CI = [87%, 103%]; p = 3.8e-6, sign test, n=19; Fig. S2F).

To measure the strength of self-inhibition, we recorded the responses of OTid neurons to a single stimulus (S1) presented inside their RF (Fig. 2G; Methods). We compared these OTid responses when Imc was intact (Fig. 2H-left) versus when the portion of Imc encoding S1, i.e., spatially matched to the OTid recording site, was focally and reversibly inactivated (Fig. 2H-right). Any difference in OTid responses between the Imc-intact (Fig. 2I-left; 2J-black) and the Imc-inactivated (Fig. 2I-right; 2J-blue) conditions quantified the amount of suppression due to Imc onto the "self" OTid location, thereby directly estimating the strength of net 'self-inhibition'.

We found that Imc neurons exerted no significant inhibition onto OTid neurons encoding for overlapping spatial locations (Fig. 2K, L-blue; mean strength = -3.7 %, s.d. = 12.2%, n = 28 neuron pairs; p = 0.12, t-test against 0; mean distance between centers = 2.86 °). We verified that these results were not due to ineffectiveness of iontophoresis by observing that the suppression of Imc responses by kynurenic acid was substantial (Fig. S2F), and not distinguishable from that in the "other" case (Fig. S2F; p = 0.68, ranksum test, Imc suppression by drug in "self" vs. "other" cases). Therefore, strength of net "self" inhibition in the OTid was substantially weaker than the strength of net competitive inhibition (Fig. 2L, red vs. blue; p = 7.8e-11, two sample t

test with HBMC correction; Methods). Together, these findings demonstrated the presence of a (net) donut-like, spatial pattern of inhibition across the OTid azimuthal space map (i.e. when S1 and S2 were separated along the azimuth).

Donut-like inhibitory motif is multi-holed

Imc neurons are strikingly asymmetric in their encoding of elevational versus azimuthal space. The majority (67%) exhibit RFs with multiple discrete firing fields ("lobes") distributed along the elevation, but not the azimuth (multilobed RFs exhibit upto 3 lobes along the elevation; (*40*); Fig. 3A,G-left panels).This unusual RF structure has been shown to be essential for Imc to achieve selection at all possible pairs of spatial locations in the face of scarcity of its neurons, and it does so does using a novel, combinatorially optimized inhibitory strategy (*40*).

A direct consequence of multilobed encoding of elevational space is that there are gaps between the lobes of an Imc RF, constituting locations that are outside that neuron's RF (Fig. 3AB-left panels; light red bar). If the donut-like inhibitory motif is to operate also for (categorical) selection among competing stimuli separated along the elevation, then the spatial pattern of inhibition must respect a very strict condition: namely, that, multilobe Imc neurons ought to send strong competitive inhibition to OTid neurons encoding locations in the gaps between RF lobes, but weak or no inhibition to OTid neurons encoding locations within the other RF lobes of the same Imc neurons (weak self-inhibition), predicting a 'multi-holed' donut (Fig. S3A).

To test experimentally if this strict requirement holds true, we again made dual extracellular recordings in the OTid and Imc. We first recorded from an Imc neuron, mapped out its spatial RF, and applied previously published analyses to determine if it was a multilobed RF

(two-lobed RF in Fig. 3AB – left panels and Fig. S3B-D; three-lobed RF in Fig. 3GH – left panels and Fig. S3E-G; Methods;(*40*)). Next, we positioned a second electrode in the OTid such that the spatial RF of the OTid neuron was either centered within the gap between Imc RF lobes (Fig. 3AB – right panels and Fig. 3C-left, light red bar) or it overlapped one of the lobes of the Imc neuron's RF (Fig. 3GH – right panels and Fig. 3I – left, light blue bar). We, then, recorded the responses of the OTid neuron to a stimulus inside its RF (S1; Fig. 3C-left and I-left; Methods) while simultaneously presenting a competing stimulus (S2) at a distant location along the elevational axis such that S2 was within a (different) lobe of the RF of the Imc (Fig. 3C-left – S1 in the Imc RF gap – light red bar and S2 within an Imc RF lobe; Fig. 3I-left – S1 within one Imc RF lobe and S2 within a different one – light blue bar). We compared OTid responses when Imc was intact (Fig. 3CI-left panels) versus when the portion of Imc encoding S2 was focally and reversibly inactivated (Fig. 3CI-right panels). As before, any observed response differences directly estimated, respectively, the strength of net inhibition exerted by the Imc neuron onto the OTid neuron encoding the gap location ("gap" inhibition), or the strength of net inhibition exerted by the Imc neuron due to a stimulus within one of its lobes, onto the OTid neuron encoding for a location within a different lobe of the Imc's RF ("different lobe" inhibition).

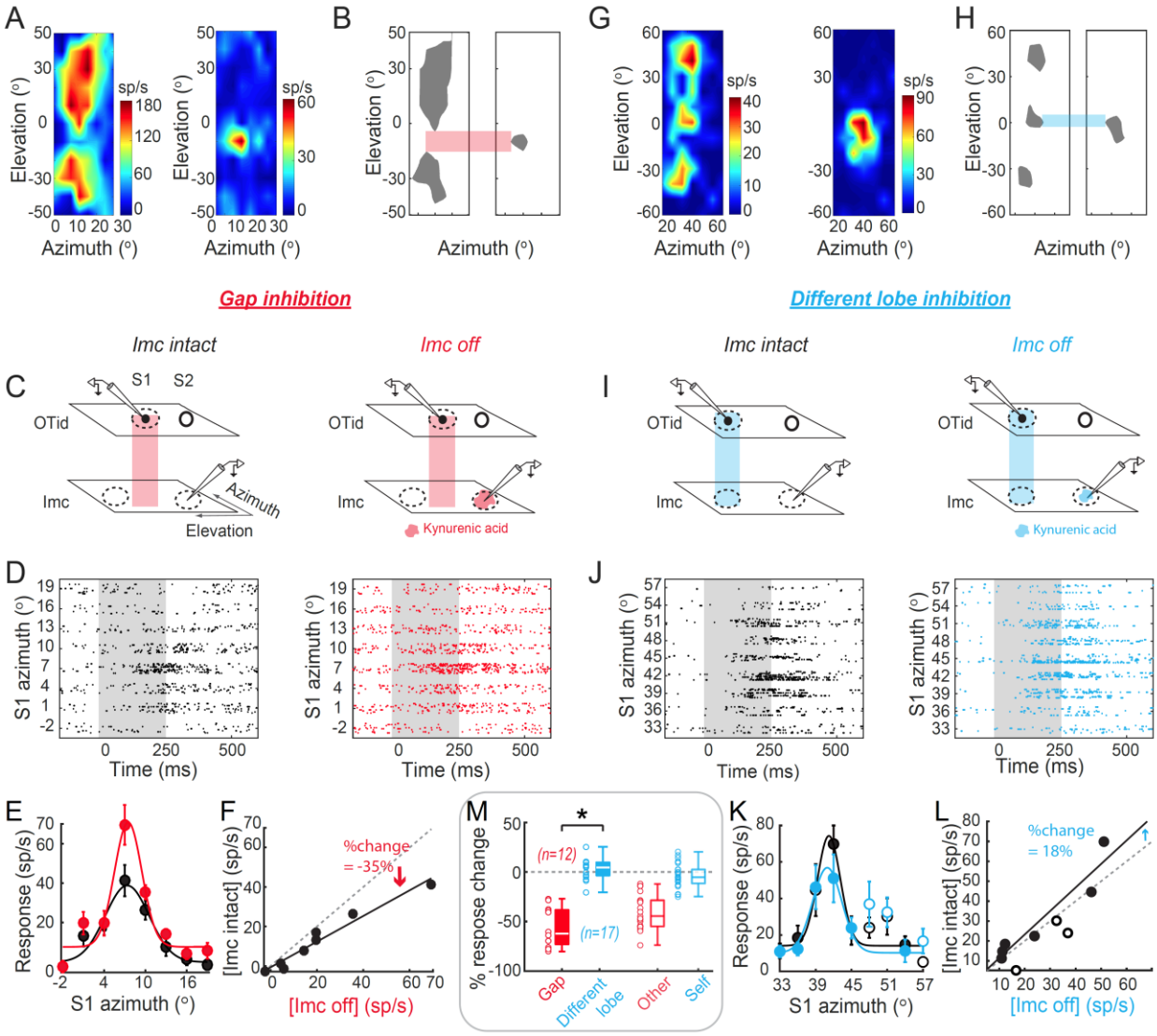


Fig. 3. Midbrain selection network in barn owl contains multi-holed donut-like inhibitory motif operating across 2-D sensory space (azimuth x elevation).

(A-F) *Measurement of the strength of net "gap" inhibition from Imc → OTid with paired recordings in Imc and OTid. (A) Spatial receptive fields (RFs) of an Imc neuron (left) and an OTid neuron (right) from a paired Imc-OTid recording experiment. Imc RF is two-lobed (text; Fig. S3B-D; Methods; (40)). OTid RF lies in the gap between the lobes of the Imc RF. (B) Binarized versions of RFs in A, at 60% max. firing rate in each case. Red horizontal bar: highlights the relative position of OTid RF to Imc RF lobes. (C-F) Conventions as in Fig. 2C-F.*

C, Red vertical bar indicates that OTid RF is in the gap between Imc RF lobes. F, Line: Best fit straight line; slope = 0.65, R^2 =0.95. % change in responses directly estimates the strength of inhibition at this OTid neuron due to Imc (here, 35%).

(G-L) *Measurement of the strength of net "different lobe" inhibition from Imc → OTid with paired recordings in Imc and OTid. Conventions as in A-F. (G, H) Three-lobed spatial RF of an Imc neuron (left; Fig. S3E-G; Methods). RF of OTid neuron overlaps with one of the lobes of Imc neuron's RF (blue bar). (L) Line: Best fit straight line; slope = 1.18, R^2= 0.9. % change in responses directly estimates the strength of inhibition at this OTid neuron due to Imc (here, -18%).*

(M) *Population summary of strength of net "gap" inhibition (red; n=12), and strength of net "different lobe" inhibition (blue; n=17) from Imc → OTid. Open red and blue box plots: reproduced from Fig. 1L. p<1e-5 (red vs. 0), p=0.35 (blue vs. 0); p=<1e-5 (red vs blue), paired t-tests with HBMC correction.*

See also Fig. S3.

We found that Imc neurons exerted strong competitive inhibition onto OTid neurons that encoded locations in the gaps between the Imc RF lobes (Fig. 3D-F, M-red), but weak competitive inhibition onto OTid neurons that encoded for locations within a different lobe of the Imc's RF (Fig. 3J-L, M-blue). The mean strength of "gap" inhibition (along elevation) was -55.88% (Fig. 3M-red, s.d. = 20%, n==12 neuron pairs; p = 1.02 e-6, t-test against 0 with HBMC correction; Fig. S3IJ-red). The mean strength of "different lobe" inhibition was 2.43 % (Fig. 3M-blue, s.d. = 10.3 %, n==17 neuron pairs; p = 0.35, t-test against 0 with HBMC correction; Fig. S3IJ-blue), and was not distinguishable from the average "self" inhibition (Fig. 3M- blue vs.

translucent blue; p-value = 0.09; two sample test with HMBC correction). Imc was inactivated effectively in both sets of experiments (median = 98.2% median, 95% CI = [96.46% 99.99%]; Fig. S3K).

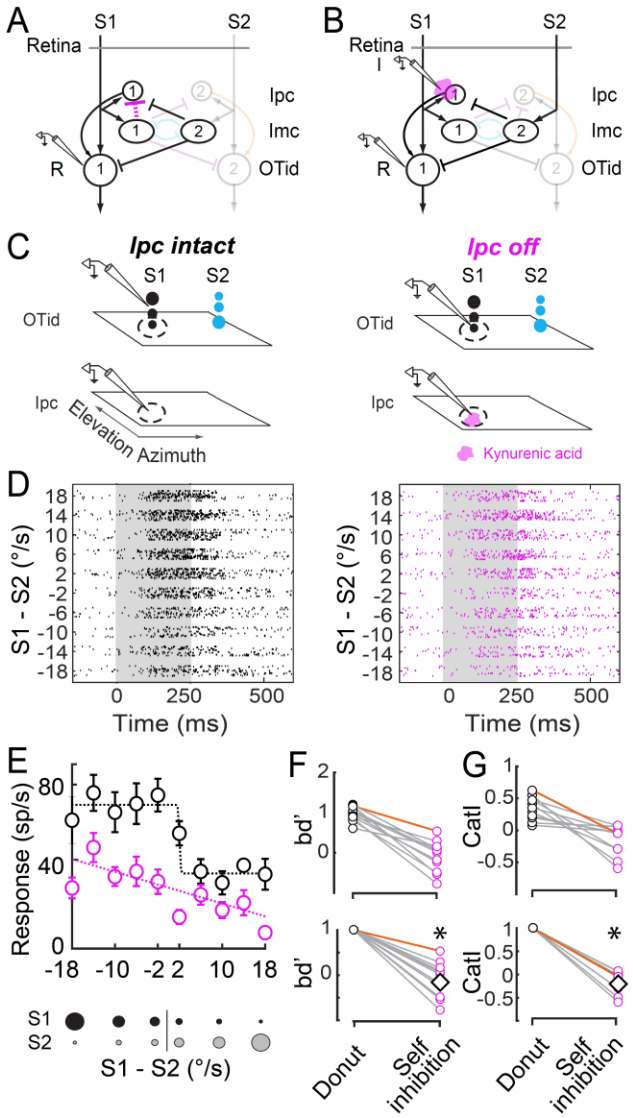
Taken together, these results demonstrated that the Imc implements precisely organized, multi-holed donut-like patterns of inhibition onto the OTid space map, with each Imc neuron's net inhibitory effect creating a pattern complementary to its (multilobed) RF structure (Fig. S3A).

Donut-like motif is required for categorization by barn owl midbrain

Considering the wiring complexity of the (multi-holed) donut-like connectivity between Imc (and Ipc) and OT, we wondered if it served a functional purpose. Specifically, we investigated whether it was necessary for the categorical signaling of the highest priority stimulus by the OTid. To test this, we reasoned that causally introducing "self" inhibition onto an OTid neuron to disrupt the donut-like pattern of inhibition would be most effective if achieved via the indirect, but potent, pathway, i.e., by increasing "self" inhibition onto Ipc, (Fig. 4A). As the net "self" inhibition at most OTid neurons is weak or zero (Figs. 2L and 3M), activating Imc → Ipc projections (or Imc → OTid projections) for spatially aligned neuron pairs would be infeasible because such projections are unlikely to exist. Instead, we mimicked this effect by focally inactivating Ipc neurons (Fig. 4B-right), while recording simultaneously the responses of spatially matched OTid neurons to the two-stimulus morphing protocol used to study categorical signaling in the OTid (Fig. 4C left vs. right; protocol shown in Fig. 4E–bottom, same as Fig. 1E–bottom; Methods;(*5*)).

Fig. 4. Donut-like inhibitory motif is required for categorical signaling of the strongest stimulus in the barn owl midbrain selection network.

(A) *Network model showing the OT-Imc-Ipc circuit in the barn owl. Magenta, dashed arrow: Introduction of "self" inhibition in the Imc→Ipc pathway; would disrupt donut-like inhibition, but is infeasible (Figs. 2 and 3; text).* ***(B)*** *Schematized is iontophoretic inactivation of Ipc neuron aligned with OTid neuron; a feasible causal manipulation that disrupts the donut-like inhibition within channel 1 by de-activating the donut-like motif onto the amplifier.* ***(C)*** *Paired recordings in OTid and Ipc such that RF of OTid neuron overlaps with that of Ipc neuron. Conventions as in Fig. 2C. Stimulus protocol used is the same two-stimulus morphing protocol used in model simulations in Fig. 1E; the relative strength between S1 and S2 was systematically varied.* ***(D)*** *Left: OTid response rasters in the Ipc-intact condition. Right: OTid response rasters in the Ipc-off condition. Distance between OTid and Ipc RF centers = 5°.* ***(E)*** *OTid response firing rates, computed from D over the 100-400 ms time window. Black: Ipc-intact condition, magenta: Ipc-off condition.* ***(F, G)*** *Population summary of effect of Ipc inactivation on bd' (F) and CatI (G); n=11 neuron pairs; data in orange are from example neuron pair in D-E. Top*



panels: measured values. Bottom panels: Data in top panels replotted after normalizing to Ipc-intact values. Diamond: average of red data; p <1e-4 (bd'), t-test, and p <1e-4 (CatI), ranksum tests against 1 with HBMC correction in F and G.

See also Fig. S4.

We found that disrupting the donut-like inhibitory motif (by introducing "self" inhibition onto the Ipc) caused a substantial reduction in categorical signaling in the OTid (Fig. 4D-E: example neuron pair)**.** Across the population of tested neuron pairs, categorization as well as robust-to-noise signaling were nearly abolished, with the median reduction in *CatI* of 104.7% and a mean reduction in *bd'* of 112.42% (Fig. 4FG; n=11 neuron pairs; *bd'*: s.d. = 37.8%; p = 3.93e-9, t-test against 1 with HBMC correction; *CatI*: 95% CI of median = [91.52%, 117.92%]; p = 2.55 e-5, ranksum test against 1 with HBMC correction). Thus, the midbrain spatial selection network not only contains the specialized donut-like inhibitory circuit motif, but also critically depends on it for categorical, robust-to-noise signaling of the highest priority stimulus.

This structured (multi-holed) donut-like organization of inhibition stands in contrast to another computational mechanism that has been invoked in the decision-making literature, namely divisive normalization (*24*, *41-43*). It involves the pooling of inhibition due to various active channels, with the inhibitory elements inhibiting each other, and with the pooled inhibition delivered *uniformly* (rather than in a donut-like manner) to the output elements (Fig. 5A;(*41*, *42*)).

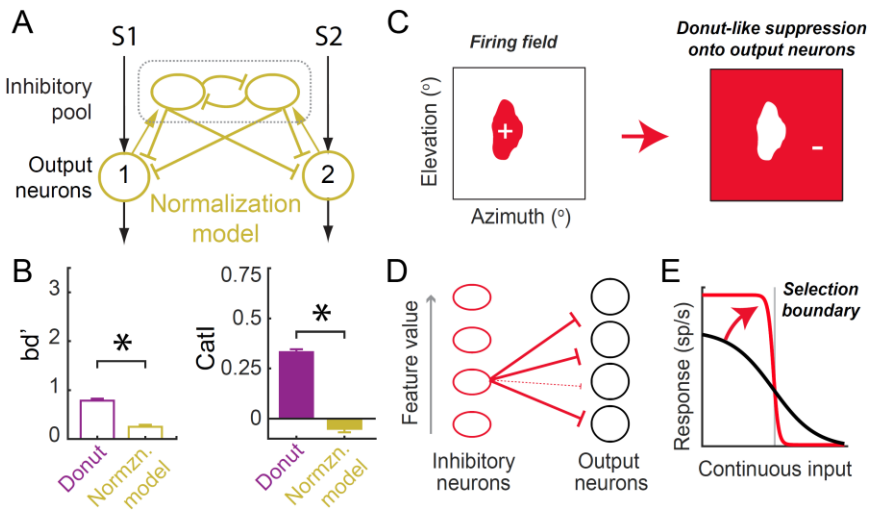


Fig. 5. Modeling: Divisive normalization model operating across the space map is not effective for producing categorical responses.

(A) *Schematic of normalization circuit model with pooled inhibition and inhibitory feedback (41, 42).* ***(B)*** *Plot of bd' (left panel; gold data) and CatI (right panel; gold data) computed from the responses of neuron 1 in above normalization model to the standard two-stimulus morphing protocol (as in Fig. 1E). For comparison, the bd' and CatI values for donut-like motif, from Fig. 1FG, reproduced here for comparison. (p = 1e-12, t-test of gold vs. purple of right panel)* ***(C-E)*** *Graphical summary. (C) Donut-like inhibition (dashed line: lacking "self" inhibition), implemented by a neural circuit (D), which sends inhibition to all non-preferred inputs, generates categorical response profiles (E), even from linear ones (Fig. 1E-inset).*

Our results above (Figs. 2-3) show that the midbrain spatial selection network (OT-Imc-Ipc) does not implement pooled divisive normalization for selection across space. To test, more generally, whether the divisive normalization model is effective for generating categorical response profiles, we simulated a circuit model of normalization (Fig. 5A) and obtained model neuron responses to the two-stimulus morphing protocol (Fig. 5A; same protocol as in Figs. 1E

and 4). *CatI* was substantially lower for this model than that obtained with just the donut-like inhibitory motif (Fig. 5B right panel; mean *CatI* = -0.05 for normalization model compared to 0.331 for donut model in Figure 1: p = 1.5 e-29, t-test, gold vs. purple; p = 1e-12, t-test of gold vs. purple, left panel). Thus, the normalization mechanism is not beneficial for generating categorical responses, consistent with findings from modeling in visual cortex (*44*). We note that, on the other hand, the donut-like motif is capable of keeping the responses of the circuit in check (in a sense, 'normalizing' the responses) as the overall drive increases (Fig. 1E).

Discussion

Our results demonstrate the power of the donut-like inhibitory motif for generating categorical representations (Fig. 5C-E), capable of converting even linear response profiles to categorical ones (Fig. 1E-inset: orange/brown vs. purple). Thus, the donut-like inhibitory motif may be a conserved neural circuit module for categorical selection and decision making in general, and may play a functional role in sculpting categorical responses that arise from learning as well (*2, 9*). Here, we described the donut-like motif in the context of selection across space, which is underpinned by well-organized topographic maps of space. However, for many forms of selection, such organized maps do not exist, with olfactory categorization being an extreme example (*4*), involving dense combinatorial coding of odors. The findings from this study suggest that detailed investigation of the properties and connectivity of inhibitory neurons in those cases may be a fruitful endeavor. In the evolutionarily conserved subcortical circuit studied here (*26, 31*), the long-range suppression that is necessary for the donut-like motif is implemented by inhibitory neurons with far reaching projections. However, it could also be implemented, for instance, in cortical circuits, through long-range excitation contacting local

inhibitory neurons. A seeming alternative mechanism for categorization that we did not discuss is one in which input-output functions of the neurons are biophysically specified (apriori) to be step-like. However, the limited power of this mechanism for producing categorical responses in general circumstances, together with the past demonstration that the steepness of the i/o functions of neurons can be uncorrelated with whether selection is categorical (*6*), made this alternative not compelling. Another alternative are highly recurrent networks which have been shown in modeling to be capable of generating categorical outputs through dynamics from multiplexed encoding (*45, 46*). However, because it is difficult to extract a specific, testable neural circuit mechanism from this solution, and in light of the recent counterpoint to such mixed-selectivity descriptions (*7*), we focused on structured circuit mechanisms. We did consider explicitly the alternatives of feedback inhibition and recurrent amplification. Contrary to prior proposals (*19, 21*), feedback inhibition, by itself, is not effective for producing categorical response profiles, nor is recurrent amplification by itself, nor the two of them combined. However, if present along with the donut-like motif, they enhance its efficacy for categorization. Our results, therefore, illuminate the potential mechanistic reason for the success of various past models that have accounted for categorical responses, namely, their inclusion of donut-like inhibition (rather than of amplification or feedback inhibition). Considering that other important functions have been proposed for these two motifs, such as implementation of a flexible selection boundary for feedback inhibition (*10, 21, 47*), and evidence accumulation for recurrent amplification (*12, 19*), circuits for selection might well possess these motifs. Together, the three motifs can effectively implement attractor dynamics for flexibly categorical decision-making (*4, 13, 19, 21*). An intriguing open question in this context, even in the avian midbrain

selection network, is how the wiring of this exquisitely organized (multi-holed) donut-like connectivity is achieved in the brain.

References and Notes:

1. X. Jiang, M. A. Chevillet, J. P. Rauschecker, M. Riesenhuber, Training Humans to Categorize Monkey Calls: Auditory Feature- and Category-Selective Neural Tuning Changes. *Neuron* **98**, 405-416 e404 (2018).
2. D. J. Freedman, J. A. Assad, Experience-dependent representation of visual categories in parietal cortex. *Nature* **443**, 85-88 (2006).
3. D. J. Freedman, M. Riesenhuber, T. Poggio, E. K. Miller, Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312-316 (2001).
4. J. Niessing, R. W. Friedrich, Olfactory pattern classification by discrete neuronal network states. *Nature* **465**, 47-52 (2010).
5. S. P. Mysore, E. I. Knudsen, Flexible categorization of relative stimulus strength by the optic tectum. *J Neurosci* **31**, 7745-7752 (2011).
6. S. P. Mysore, A. Asadollahi, E. I. Knudsen, Signaling of the strongest stimulus in the owl optic tectum. *J Neurosci* **31**, 5186-5196 (2011).
7. J. Hirokawa, A. Vaughan, P. Masset, T. Ott, A. Kepecs, Frontal cortex neuron types categorically encode single decision variables. *Nature* **576**, 446-451 (2019).
8. B. Bathellier, L. Ushakova, S. Rumpel, Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron* **76**, 435-449 (2012).
9. Y. Xin *et al.*, Sensory-to-Category Transformation via Dynamic Reorganization of Ensemble Structures in Mouse Auditory Cortex. *Neuron* **103**, 909-921 e906 (2019).
10. T. Jovanic *et al.*, Competitive Disinhibition Mediates Behavioral Choice and Sequences in Drosophila. *Cell* **167**, 858-870 e819 (2016).
11. T. A. Engel, W. Chaisangmongkon, D. J. Freedman, X. J. Wang, Choice-correlated activity fluctuations underlie learning of neuronal category representation. *Nature communications* **6**, 6454 (2015).
12. X. J. Wang, Neural dynamics and circuit mechanisms of decision-making. *Curr Opin Neurobiol* **22**, 1039-1046 (2012).
13. C. K. Machens, R. Romo, C. D. Brody, Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* **307**, 1121-1124 (2005).
14. S. P. Mysore, N. B. Kothari, Mechanisms of competitive selection: A canonical neural circuit framework *In review*
15. S. P. Mysore, A. Asadollahi, E. I. Knudsen, Global inhibition and stimulus competition in the owl optic tectum. *J Neurosci* **30**, 1727-1738 (2010).
16. S. R. Olsen, V. Bhandawat, R. I. Wilson, Divisive normalization in olfactory population codes. *Neuron* **66**, 287-299 (2010).
17. A. Bollimunta, J. Ditterich, Local computation of decision-relevant net sensory evidence in parietal cortex. *Cereb Cortex* **22**, 903-917 (2012).

18. H. K. Hartline, H. G. Wagner, F. Ratliff, Inhibition in the eye of Limulus. *J Gen Physiol* **39**, 651-673 (1956).
19. X. J. Wang, Decision making in recurrent neuronal circuits. *Neuron* **60**, 215-234 (2008).
20. Q. Chen, Z. Pei, D. Koren, W. Wei, Stimulus-dependent recruitment of lateral inhibition underlies retinal direction selectivity. *eLife* **5**, (2016).
21. S. P. Mysore, E. I. Knudsen, Reciprocal inhibition of inhibition: a circuit motif for flexible categorization in stimulus selection. *Neuron* **73**, 193-205 (2012).
22. M. I. Sereno, P. S. Ulinski, Caudal topographic nucleus isthmi and the rostral nontopographic nucleus isthmi in the turtle, Pseudemys scripta. *J Comp Neurol* **261**, 319-346 (1987).
23. X. J. Wang, Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36**, 955-968 (2002).
24. L. Albantakis, G. Deco, The encoding of alternatives in multiple-choice decision making. *Proc Natl Acad Sci U S A* **106**, 10308-10313 (2009).
25. S. P. Mysore, E. I. Knudsen, The role of a midbrain network in competitive stimulus selection. *Curr Opin Neurobiol* **21**, 653-660 (2011).
26. E. I. Knudsen, Control from below: the role of a midbrain network in spatial attention. *Eur J Neurosci* **33**, 1961-1972 (2011).
27. L. P. Lovejoy, R. J. Krauzlis, Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. *Nat Neurosci* **13**, 261-266 (2010).
28. R. M. McPeek, E. L. Keller, Deficits in saccade target selection after inactivation of superior colliculus. *Nat Neurosci* **7**, 757-763 (2004).
29. J. H. Fecteau, D. P. Munoz, Salience, relevance, and firing: a priority map for target selection. *Trends Cogn Sci* **10**, 382-390 (2006).
30. S. P. Mysore, E. I. Knudsen, Descending control of neural bias and selectivity in a spatial attention network: rules and mechanisms. *Neuron* **84**, 214-226 (2014).
31. Y. Wang, D. E. Major, H. J. Karten, Morphology and connections of nucleus isthmi pars magnocellularis in chicks (Gallus gallus). *J Comp Neurol* **469**, 275-297 (2004).
32. S. P. Mysore, E. I. Knudsen, A shared inhibitory circuit for both exogenous and endogenous control of stimulus selection. *Nat Neurosci* **16**, 473-478 (2013).
33. G. Marin *et al.*, A cholinergic gating mechanism controlled by competitive interactions in the optic tectum of the pigeon. *J Neurosci* **27**, 8112-8121 (2007).
34. E. I. Knudsen, Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.* **2**, 1177-1194 (1982).
35. C. A. Goddard, S. P. Mysore, A. S. Bryant, J. R. Huguenard, E. I. Knudsen, Spatially Reciprocal Inhibition of Inhibition within a Stimulus Selection Network in the Avian Midbrain. *PLoS One* **9**, e85865 (2014).
36. Y. Wang, H. Luksch, N. C. Brecha, H. J. Karten, Columnar projections from the cholinergic nucleus isthmi to the optic tectum in chicks (Gallus gallus): a possible substrate for synchronizing tectal channels. *J Comp Neurol* **494**, 7-35 (2006).
37. G. Marin, J. Mpodozis, E. Sentis, T. Ossandon, J. C. Letelier, Oscillatory bursts in the optic tectum of birds represent re-entrant signals from the nucleus isthmi pars parvocellularis. *J Neurosci* **25**, 7081-7089 (2005).
38. A. Asadollahi, E. I. Knudsen, Spatially precise visual gain control mediated by a cholinergic circuit in the midbrain attention network. *Nature communications* **7**, 13472 (2016).

39. G. Rizzolatti, R. Camarda, L. A. Grupp, M. Pisa, Inhibitory effect of remote visual stimuli on visual responses of cat superior colliculus: spatial and temporal factors. *J Neurophysiol* **37**, 1262-1275 (1974).
40. N. R. Mahajan, S. P. Mysore, Combinatorial Neural Inhibition for Stimulus Selection across Space. *Cell reports* **25**, 1158-1170 e1159 (2018).
41. M. Carandini, D. J. Heeger, J. A. Movshon, Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* **17**, 8621-8644 (1997).
42. M. Kouh, T. Poggio, A canonical neural circuit for cortical nonlinear operations. *Neural Comput* **20**, 1427-1451 (2008).
43. K. Louie, M. W. Khaw, P. W. Glimcher, Normalization is a general neural mechanism for context-dependent decision making. *Proc Natl Acad Sci U S A* **110**, 6139-6144 (2013).
44. L. Busse, A. R. Wade, M. Carandini, Representation of concurrent stimuli by population activity in visual cortex. *Neuron* **64**, 931-942 (2009).
45. V. Mante, D. Sussillo, K. V. Shenoy, W. T. Newsome, Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78-84 (2013).
46. W. Chaisangmongkon, S. K. Swaminathan, D. J. Freedman, X. J. Wang, Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential, Category-Based Decisions. *Neuron* **93**, 1504-1517 e1504 (2017).
47. J. P. Fadok *et al.*, A competitive inhibitory circuit for selection of active and passive fear responses. *Nature* **542**, 96-100 (2017).
48. A. Asadollahi, S. P. Mysore, E. I. Knudsen, Stimulus-driven competition in a cholinergic midbrain nucleus. *Nat Neurosci* **13**, 889-895 (2010).
49. F. J. Chaure, H. G. Rey, R. Quian Quiroga, A novel and fully automatic spike-sorting implementation with variable number of features. *J Neurophysiol* **120**, 1859-1871 (2018).
50. A. Asadollahi, S. P. Mysore, E. I. Knudsen, Rules of competitive stimulus selection in a cholinergic isthmic nucleus of the owl midbrain. *J Neurosci* **31**, 6088-6097 (2011).
51. A. Rodriguez, A. Laio, Machine learning. Clustering by fast search and find of density peaks. *Science* **344**, 1492-1496 (2014).
52. R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a dataset via the gap statistic. *J. R. Statist. Soc. B* **63**, 411-423 (2001).

Acknowledgments:

Funding: This work was supported in part by funding from NIH R01 EY027718. **Author contributions:** NRM and SPM designed the research, NRM performed experiments and analyzed the data, SPM and NRM wrote the paper. **Competing interests:** Authors declare no competing interests.

Supplementary Materials:

Materials and Methods

Figures S1-S4

References (*48-52*)

Supplementary Materials for

Neural circuit mechanism for generating categorical representations

Materials and Methods

Animals

We performed experimental recordings in 7 head-fixed awake adult barn owls viewing a visual screen passively (Tyto alba). Both male and female birds were used; the birds were shared across studies. All procedures for animal care and use were carried out following approval by the Johns Hopkins University Institutional Animal Care and Use Committee, and in accordance with NIH guidelines for the care and use of laboratory animals. Owls were group housed in flight runs within the aviary, each containing up to 6 birds. The light/dark cycle was 12 hr/12 hr.

Neurophysiology

Experiments were performed following protocols that have been described previously (*15, 32*). Briefly, epoxy-coated, high impedance, tungsten microelectrodes (A-M Systems, 250μm, 5-10MΩ at 1 kHz) were used to record single and multi-units extracellularly in the OTid. Multi-barrel glass electrodes (Kation Scientific, Carbostar– 3LT, 0.4-1.2MΩ at 1kHz) filled with kynurenic acid (a competitive inhibitor of ionotropic glutamate receptors; pH 8.5-9 at a concentration of 40mM) were used to record from and inactivate neurons in the Imc and Ipc. Inactivation was performed using micro iontophoresis by ejecting kynurenic acid with an eject current of -450 nA to -500 nA; data were collected starting 15 min after drug ejection commenced. A retain current of +15 nA was used to prevent leakage of the drug from the tip of the electrodes when drug was not being

iontophoresed. Recovery data were measured 15 min after drug ejection was ceased.

OT, Imc and Ipc targeting

We navigated to the OT (based on well-established methods (*34*)), and then navigated to the Imc using the OT's topographic space map as reference. The Imc is an oblong structure that is 2.8mm rostrocaudally and 0.35mm dorsoventrally, appearing as a 700-µm x 350-µm elliptical disk in coronal sections. It lies parallel to the rostrocaudal axis of the OT, located approximately 16 mm ventral to the surface of the brain, and approximately 500 µm medial to the medial-most part of the OT. We targeted the Imc following published methods (*32*) Imc targeting has been validated previously using dye injections and lesions (*40*). Dorsoventral penetrations through the Imc were made at a medial-leading angle of 5˚ from the vertical to avoid a major blood vessel in the path to the Imc. The Ipc lies roughly 500-700 um medial to the Imc and its targeting was confirmed based on the neural response characteristics of the neurons (characteristic bursty responses; (*48*)).

Data collection and spike sorting

Multi-unit spike waveforms were recorded using Tucker Davis Technologies hardware interfaced with MATLAB. The responses of neurons were measured by counting spikes during a time window (typically 100-350 ms) following stimulus onset.

The automated 'wave-clus' spike-sorting toolbox was used for spike sorting(*49*). We included only those units for analysis for which fewer than 5% of the recorded spikes were within 1.5 ms (inter-spike interval; ISI) of each other.

Model details (Related to Figs. 1, S1, S2, 5)

Input output functions

The input output functions (firing rate f, as a function of the saliency, l) of the neurons in the model were modeled using sigmoid functions using previously published methods(*21*).

$$f(l) = c + s\left(\frac{l^m}{l^m + L_{50}^m}\right)$$

where c, is the baseline firing rate of the neuron; l, is the saliency parameter of the stimulus (e.g. loom speed of the stimulus, loudness of an auditory stimulus, contrast of the stimulus, speed of a moving stimulus) that can vary continuously over a range; s, is the maximum change in the firing rate of the neuron; L_{50} is the saliency value at which the neuron's firing rate changes by 50% of the maximum change; and m, is a parameter that controls the slope of the sigmoid.

Excitatory neurons: The excitatory neurons in Figs. 1, S1, S2, and S4 were simulated using the following parameters (which are consistent with the parameters obtained by fitting a sigmoid to response functions of OTid neurons (*21*):

$$c = 5.3, s = 22.2\ , L_{50} = 11.6, m = \ 2$$

Inhibitory neurons: The inhibitory neurons in Figs. 1, S1, S2, and S4 were simulated using the following parameters (which are consistent with the parameters obtained by fitting a sigmoid to response functions of Imc neurons (*21*):

$$c = 5, s = 15\ , L_{50} = 8, m = 10$$

<u>*Recurrent excitation neurons*</u>: The excitatory neurons that provide recurrent amplification in Fig. S2 were simulated using the following parameters (which are consistent with the parameters obtained by fitting a sigmoid to response functions of Ipc neurons; (*50*)):

$$c = 8.4, s = 36\ , L_{50} = 5.8, m = \ 3.3$$

The inhibition sent from the inhibitory neurons onto excitatory neurons is modeled using input and output divisive factors as below using previously published methods (*21*).

$$f = \left(\frac{1}{s_{out}} + 1\right).\left(\frac{c}{s_{in}+1} + s\left(\frac{l^m}{l^m + L_{50}^m + s_{in}^m}\right)\right), \text{ where}$$

- $s_{in} = d_{in}.I, s_{out} = d_{out}.I$ are the input and output divisive factors.

- $d_{in}\ and\ d_{out}$ are parameters that control the strength of input and output division and I is the output (firing rate) of the inhibitory neuron sending the inhibition.

The value of these parameters chosen were $d_{in} = 0\ and\ d_{out} = 0.25$ consistent with previously published methods (see Fig. 5D in (*21*)).

Feedback inhibition

In models in which the inhibitory neurons inhibit each other (e.g. Fig. 1C, bottom left panel, blue connections; feedback model), the feedback inhibition was modeled as below using previously published methods(*21*).

$$I(t) = \left(\frac{1}{i_{out}} + 1\right).\left(\frac{c}{i_{in} + 1} + s\left(\frac{l^m}{l^m + L_{50}^m + i_{in}^m}\right)\right)$$

$$i_{in}(t) = r_{in}.I'(t-1), \quad i_{out} = r_{out}.I'(t-1)$$

Where I' is the output (firing rate) of the other inhibitory neuron at time $(t-1)$.

These equations were iteratively applied until there was no further change in the output of the neurons (i.e., steady state was reached).

The values of the feedback parameters used were $r_{in} = 0.8, r_{out} = 0.01$ consistent with previously published methods. We also varied these two parameters (varying feedback; Fig. S1A) to study their effect on boundary discriminability and categorization index.

"Self" inhibition and Donut

In models in which the excitatory neuron receives inhibition from more than one inhibitory neuron (e.g. Fig. 1C, top left panel; Baseline model), the inhibition from these sources was combined as below.

$$f = \left(\frac{1}{s_{out1}} + 1\right).\left(\frac{1}{s_{out2}} + 1\right).\left(\frac{c}{s_{in1} + s_{in2} + 1} + s\left(\frac{l^m}{l^m + L_{50}^m + s_{in1}^m + s_{in2}^m}\right)\right)$$

where, $s_{in1} = d_{in}.I_1, s_{out1} = d_{out}.I_1, s_{in2} = (s * d_{in}).I_2, s_{out2} = (s * d_{out}).I_2$

I_1 is the output (firing rate) of the inhibitory neuron 1 and I_2 is the output (firing rate) of the inhibitory neuron 2.

The parameter value s controls the strength of self-inhibition, and ranges between 0 and 1. In models that have 'donut-like connectivity', s is set to 0 (Fig. 1C, top right panel; Donut model). In models which do not have the donut-like connectivity (Fig. 1C, top left panel; Baseline model), s is set to 1.

We also vary the value of s systematically between 0 (donut) and 1 (maximum self-inhibition) to study the effect of the strength of "self" inhibition on the boundary discriminability and categorization index (Fig. S1B).

Recurrent excitation

In the models with recurrent excitation, (Fig. 1D), the output of the neuron is scaled by a factor (k, k= 2.5 in Fig. 1FG). We also vary the scaling factor to study its effect on boundary discriminability and categorization index (Fig. S1C).

In the model in Fig. S2, the output of the neuron which receives recurrent amplification is modeled as below.

$$f(l) = (1 + e_{out}).\left(\frac{1}{s_{out} + 1}\right).(\frac{c}{s_{in} + 1} + s\left(\frac{l^m}{l^m + L_{50}^m + s_{in}^m}\right))$$

where, $e_{out} = e_a.A, \quad e_a = 0.01$

and A is the output of the neuron sending the amplification.

The value of the parameter is chosen as 0.01 to yield results consistent with the amplification effect of Ipc on OT responses as reported in previously published work (Ipc inactivation results in a 31% decrease in the OTid responses on average; (*38*)).

Normalization model: Fig. 5

To implement the normalization model, we used input-output functions and the effect of divisive inhibition as described above. The primary feature of the normalization model different from models in Figs. 1, S1 and S2 is that the excitatory neurons send inputs to inhibitory neurons which then send pooled inhibition to all excitatory neurons. Neurons in the inhibitory pool inhibit each other (in addition to the feedback loop between the inhibitory neurons). The output of an excitatory neuron at time t is calculated by applying divisive inhibition (from the inhibitory pool at time t-1) to its activity at time t-1. This activity is then used to calculate the activity of the neurons in the inhibitory pool at time t, and the process is iteratively repeated. We ran this model until steady

state activity was obtained.

Calculating boundary discriminability and categorization index for the models (Figs. 1, S1, S2, 5)

For each of the models in Fig 1, S1 and S2, we used a two-stimulus strength morphing protocol described in previously published work (*5*). We presented stimulus S1 at location 1 and stimulus S2 at location 2. As the strength of the first stimulus was increased, the strength of the second stimulus was decreased (Figs. 1E, 4E). Responses of the model output neuron (#1) were simulated using this protocol. Random noise from a standard normal distribution was added to the responses. The variance of the noise added depended on the mean value of the responses (m) and a fano factor value (ff): variance = ff *m. This was repeated 30 times (to mimic 30 reps of data collection from a 'neuron') and was used to compute the response profiles (mean +/- s.e.m.) of that neuron from different model circuits. Similarly, response profiles for 50 neurons were obtained for each model circuit. For all the model runs reported in Figs. 1, S1, S2 and 5, we used a fano factor value of 6. We also varied the fano factor value to test the effect of noise on model performance (Fig. S1D).

This stimulus presentation protocol results in 2 categories (Category 1: S1 > S2 and Category 2: S1 < S2). We measured the boundary discriminability (bd') of the responses of neuron 1 by calculating the d-prime between the responses of neuron to stimuli pair straddling either side of the selection boundary and at a distance of 3 units from the boundary as:

$$bd' = \frac{\mu_1 - \mu_2}{\sqrt{0.5\left(s_1^2 + s_2^2\right)}}$$

where, μ_1 and $s_1$are the mean and the standard deviation of the responses of the neuron to the stimuli pair from category 1 near the boundary; and μ_2 and $s_2$are the mean and the standard deviation of the responses of the neuron to the stimuli pair from category 2 near the boundary.

To compute the categorization index, we compared two quantities: (modified from previously published work; (*2, 5*)) (a) the mean within-category d-prime (WCD') between the responses of the neuron to pairs of stimulus-pairs (S1 and S2) that are in the same category, and (b) the mean between-category d-prime (BCD') between the responses of the neuron to pairs of stimulus-pairs that are in different categories. The pairs are chosen while ensuring that (i) the number of pairs used to calculate both these metrics are the same, and (ii) the distribution of the distances between the chosen pairs for calculating both the metrics are the same. The categorization index is calculated from these two metrics as:

$$\text{Categorization index (CatI)} = \frac{\text{mean}(\text{BCD}') - \text{mean}(\text{WCD}')}{\text{mean}(\text{BCD}') + \text{mean}(\text{WCD}')}$$

CatI= 1 indicates idealized, step-like responses; = 0 indicates linear, non-categorical responses; <0 indicates better discriminability within category than between categories.

<u>Data collection protocol:</u>

Visual stimuli used here have been described previously (*5, 15*). Briefly, looming visual dots are flashed at different locations on a tangent TV monitor in front of the owl. Looming stimuli were dots that expanded linearly in size over time, starting from a size of 0.6˚ in radius. The speed of the loom was decided based on the stimulus protocol and varied between 9.6°/s and 21.6°/s. Visual

stimuli were presented for a duration of 250ms with an inter stimulus interval of 1.25s to 4s.

Receptive fields (RFs): For measuring spatial RFs of Imc, OTid and Ipc neurons, a single stimulus of a fixed contrast was presented at the sampled locations. The order of locations at which the stimulus was presented was randomized to minimize adaptation. A location is considered to be inside the RF if it evokes responses significantly different from that of baseline response. All other locations are considered to be outside the RF.

The following stimulus protocols were used for measuring data reported in Figs. 2, 3.

I. Paired OTid and Imc data collection

1. *"Other" inhibition*

To measure the strength of "other" inhibition sent from an Imc neuron to a distant location in the OTid space map, we simultaneously recorded from the Imc neuron (with a multi-barrel glass electrode filled with the kynurenic acid), and a spatially misaligned "other" OTid neuron (with a tungsten electrode) as described below; we ensured that the half-max of the RF of the OTid neuron lay outside (did not overlap with) the half-max of the Imc RF.

For measuring "other" inhibition, we recorded the following data curves together in an interleaved manner.

a) Tuning curve (TC) centered at the OTid RF peak: A 1-dimensional spatial tuning curve centered at the peak of the OTid RF. The stimulus had a loom strength of 9.6°/s.

b) Tuning curve centered at the OTid RF peak + competing stimulus centered at the Imc RF peak (TCC): The same curve as in a), but along with a (distant) competitor positioned at the peak of the Imc RF. The strength of the competing stimulus was chosen to be between 19.2°/s. This second competing stimulus drives the Imc neuron, which sends strong competitive inhibition to the OTid neuron. Since the competing stimulus was more salient than the stimulus driving the OTid neuron, the responses of the OTid neuron were strongly suppressed consistent with previous published results (*15*).

c) Tuning curve (TC) centered at the Imc RF peak: 1-dimensional spatial tuning curve centered at the Imc RF. The loom strength of the stimulus was chosen to be 9.6°/s.

We measured the above 3 curves both when the Imc neuron is intact ('baseline' condition) and focally inactivated ('inactivation' condition) and compare the responses as below. In a subset of the data, we also measured the curves after the responses of the Imc neurons recovered (Fig. S2D, 'recovery' condition) from focal iontophoretic inactivation. Inactivation curves were measured 15 min after the drug ejection was started. Recovery curves were measured 15 min after the drug ejection was stopped.

We analyzed the data from these three curves as below.

First, a distant competing stimulus is known to typically suppress OTid responses (*15*, *39*), something that we confirmed by comparing the responses of OTid neurons to the TC (curve a)) and TCC (curve b)); Fig. S2E. Notably, it is also known that some OTid neurons do not show such suppression, and any such neurons were excluded from our analyses (*6*).

Next, to measure the amount of Imc inactivation, we compared the responses of the Imc neurons to the TC measured at the Imc RF peak (curve c) in the baseline condition versus the inactivation condition. Kynurenic acid was able to effectively shut down Imc responses: (Fig 2F; red, median strength of Imc inactivation = 95%, 95% CI = [87%, 103%], p = 3.8e-6, sign test, n=19).

Finally, to measure the strength of "other" inhibition, we compared the TCC responses (curve b) in the baseline condition and the inactivation condition. In the baseline condition, the Imc neuron was intact, driven by the competing stimulus and sent strong inhibition to the OTid neuron. In the inactivation condition, the Imc neuron was shut off, as the result of which the OTid neuron was released from inhibition and exhibited an increase in responses. The difference between these two curves quantified the strength of "other" inhibition from the Imc onto OTid.

2. *"Self" inhibition*:

To measure the strength of "self" inhibition sent from an Imc neuron to a matched location in the OTid space map, we simultaneously recorded from the Imc neuron (with a multi-barrel glass electrode filled with the kynurenic acid), and a spatially aligned "self" OTid neuron (with a tungsten electrode) as described below; we ensured that the half-max of the RF of the OTid neuron overlapped with the half-max of the Imc RF.

We recorded a spatial tuning curve centered at the peak of the OTid (and therefore, Imc) RF. The loom speed of the stimulus used was 9.6°/s; this stimulus drives both the OTid and the Imc neuron.

We compared the responses of the OTid neuron in the Imc-intact (baseline) and Imc inactivated condition and the difference quantified the strength of "self" inhibition (Fig. 2L, blue).

To measure the amount of Imc inactivation, we compared the responses of the Imc neuron in the baseline and the inactivation condition as before. kynurenic acid was able to effectively shut down Imc responses: (median strength of Imc inactivation = 92%, 95% CI = [86% 98%], p = 7.5e-9, sign test).

3. *"Gap" inhibition*:

To measure the strength of "self" inhibition sent from an Imc neuron with a multilobed RF to the locations in the OTid space map between the RF lobes (gaps), we simultaneously recorded from the Imc neuron (with a multi-barrel glass electrode filled with the kynurenic acid), and "gap" OTid neuron (with a tungsten electrode), the RF (half-max) of which was located in the gap between the half-max of the lobes of the Imc RF (Fig. 3A-C).

We recorded the same set of curves as in the "other" inhibition case (see above) from the gap OTid neuron, and performed similar analyses on the data to quantify the strength of "gap" inhibition.

4. *"Different lobe" inhibition*:

To measure the strength of inhibition sent “from” locations within one lobe of an Imc neuron with a multilobed RF to the locations in the OTid space map within other lobes of the Imc RF (“different lobe” inhibition), we simultaneously recorded from a multilobe Imc neuron (with a multi-barrel glass electrode filled with the kynurenic acid), and an OTid neuron (with a tungsten electrode), the RF (half-max) of which overlapped with one of the lobes of the Imc RF (Fig. 2G-I). Then we measured the same set of curves as in the “other” case while ensuring one additional detail. We recorded the TC curve (curve a) with stimulus (S1) centered at the OTid RF peak (and therefore within one of the lobes of the Imc RF as well). For the TCC (curve b), the second stimulus S2 was centered at the peak of a different lobe of the Imc neuron’s multilobe RF. Thus, both stimuli excited the Imc neuron (during curve b) as they lay within its RF, but only S1 excited the OTid neuron and S2 served as a (distant) competitor from the its perspective.

An OTid neuron in this configuration was defined as a “different lobe” OTid neuron; the half max of its RF overlapped with the half max of one of the lobes of the Imc RF, but not of other lobes. (Note that by definition, every “different lobe” OTid neuron is also a potentially “self” OTid neuron, and admits to the use of the appropriate stimulus protocol to measure “self” inhibition. However, the opposite is not true, since “self” OTid neurons can be identified for Imc neurons with single lobed RFs as well, but “different lobe” OTid neurons are only defined for multilobe Imc neurons).

We applied similar analyses as in the “other” case to the data from “different lobe” OTid neurons and quantified the strength of “different lobe” inhibition.

II. Paired OTid and Ipc data collection

To test if the donut-like inhibitory motif is required for (robustness-to-noise and) categorization, we made paired recordings at spatially aligned Ipc and OTid neurons. We used a strength morphing stimulus protocol described in previously published work (*5*). Briefly, we presented one stimulus (S1) inside the RF of the OTid neuron (and also Ipc neuron because their RFs overlap). Simultaneously, we presented a competing stimulus (S2) 30˚ away along azimuth from S1. As the strength of the S1 decreased, the strength of S2 stimulus increased (Fig. 4E).

We applied the same analyses described above for the model in Figs. 1, S1 to compute the boundary discriminability and the categorization index for experimental data reported in Fig. 4.

Statistical Analysis

All analyses were carried out with custom MATLAB code. Parametric or non-parametric statistical tests were applied based on whether the distributions being compared were Gaussian or not, respectively (Lilliefors test of normality). The Holm-Bonferroni correction was used to account for multiple comparisons; the abbreviation "with HBMC correction" in the text stands for "with Holm-Bonferroni correction for multiple comparisons". All tests were two-sided. Data shown as a ± b refer to mean ± standard deviation, unless specified otherwise. The '*' symbol indicates significance at the 0.05 level (after corrections for multiple comparisons, if applicable). Correlations were tested using Pearson's correlation coefficient (*corr* command in MATLAB with the 'Pearson' option).

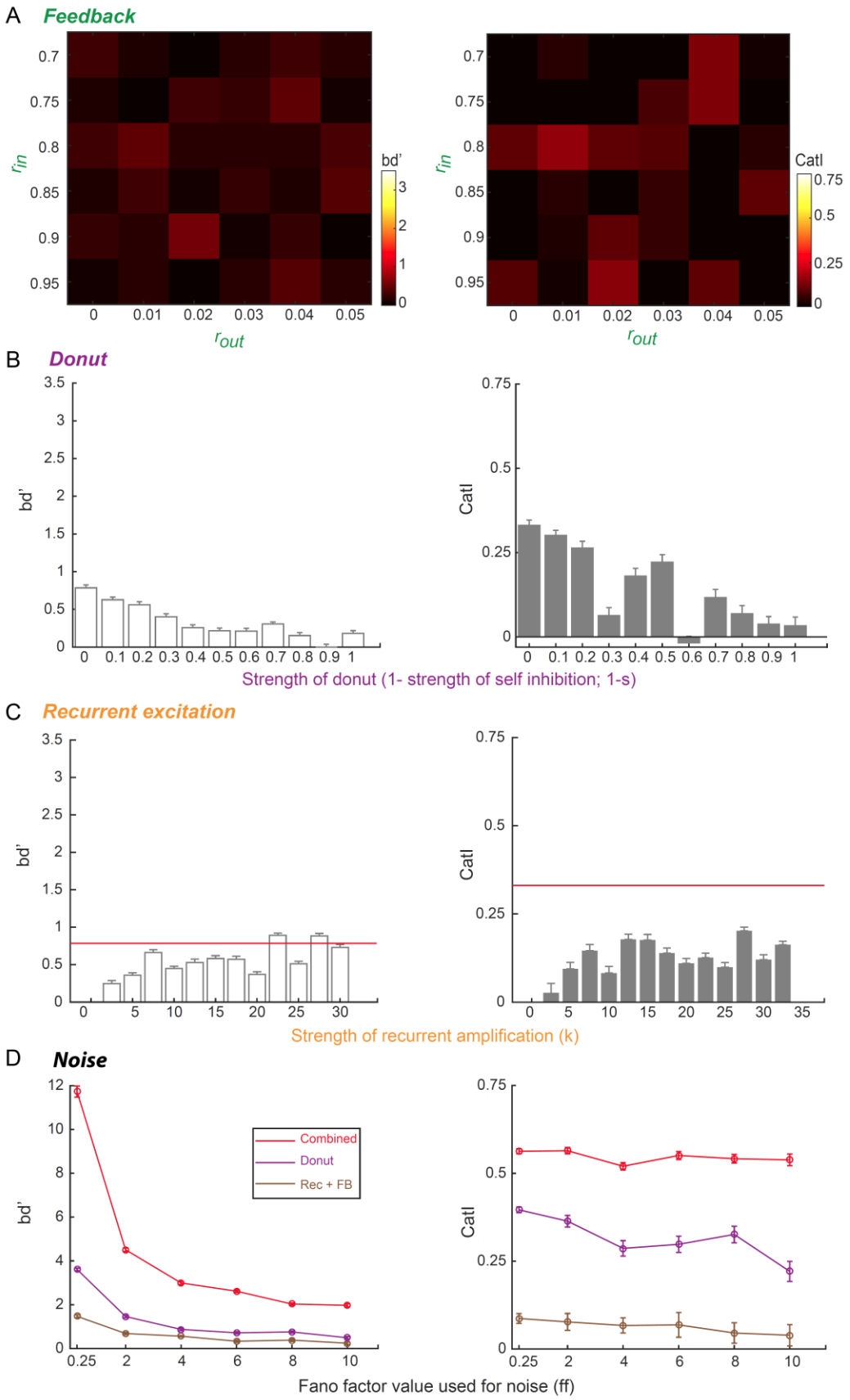

A Feedback
bd'
CatI
r_{in}
r_{out}
B Donut
bd'
CatI
Strength of donut (1- strength of self inhibition; 1-s)
C Recurrent excitation
bd'
CatI
Strength of recurrent amplification (k)
D Noise
Combined
Donut
Rec + FB
bd'
CatI
Fano factor value used for noise (ff)

Fig. S1. Donut-like inhibition surpasses other circuit motifs in its ability to generate categorical representations over a large range of values of key model parameters and amounts of response noise (modeling).

(A) *Effect of varying strength of feedback inhibition in model in Fig. 1C-bottom-left, on bd' (left) and CatI (right).* r_{in}*: input divisive factor;* r_{out}*: output divisive factor (Methods). Range of variation based on previously published work (21). Maximum bd' over the range of parameters is = 0.5478 (lower than from the donut-like motif, bd' = 0.785); maximum CatI = 0.157 (lower than from the donut-like motif, CatI = 0.331).* ***(B)*** *Effect of varying strength of "self" inhibition in model in Fig. 1C-top-right on bd' (left) and CatI (right). Left: Corr = -0.89, p = 1.64e-4 (correlation test). Right: Corr = -0.81, p = 3e-3 (correlation test). Maximal effects on bd' and CatI are when "self" inhibition=0, i.e., when the circuit has donut-like inhibition.* ***(C)*** *Effect of varying strength of recurrent amplification in model in Fig. 1D-top-left, on bd' (left) and CatI (right). Red lines: values from circuit with donut-like motif only. Left: bd' increases with amplification (corr=0.68; p = 1e-3; corr test), saturating around k=10; the increase is as expected because of the progressive scaling up of responses due to amplification. Note that physiological range is under 3 (38). Right: However, CatI does not change systematically (corr= -0.14; p =0.54; corr test), indicating that varying the strength of amplification either does not affect the estimates in Fig. 1F for this motif.* ***(D)*** *Comparison across three models of bd' (left) and CatI (right) from simulated response profiles as a function of fano-factor. Models are circuit with donut-like inhibitory motif only (Fig. 1C, top-right), or circuit with recurrent amplification and feedback inhibition together (Fig. 1D, bottom-left), or circuit with all three motifs combined (Fig. 1D, bottom-right).*

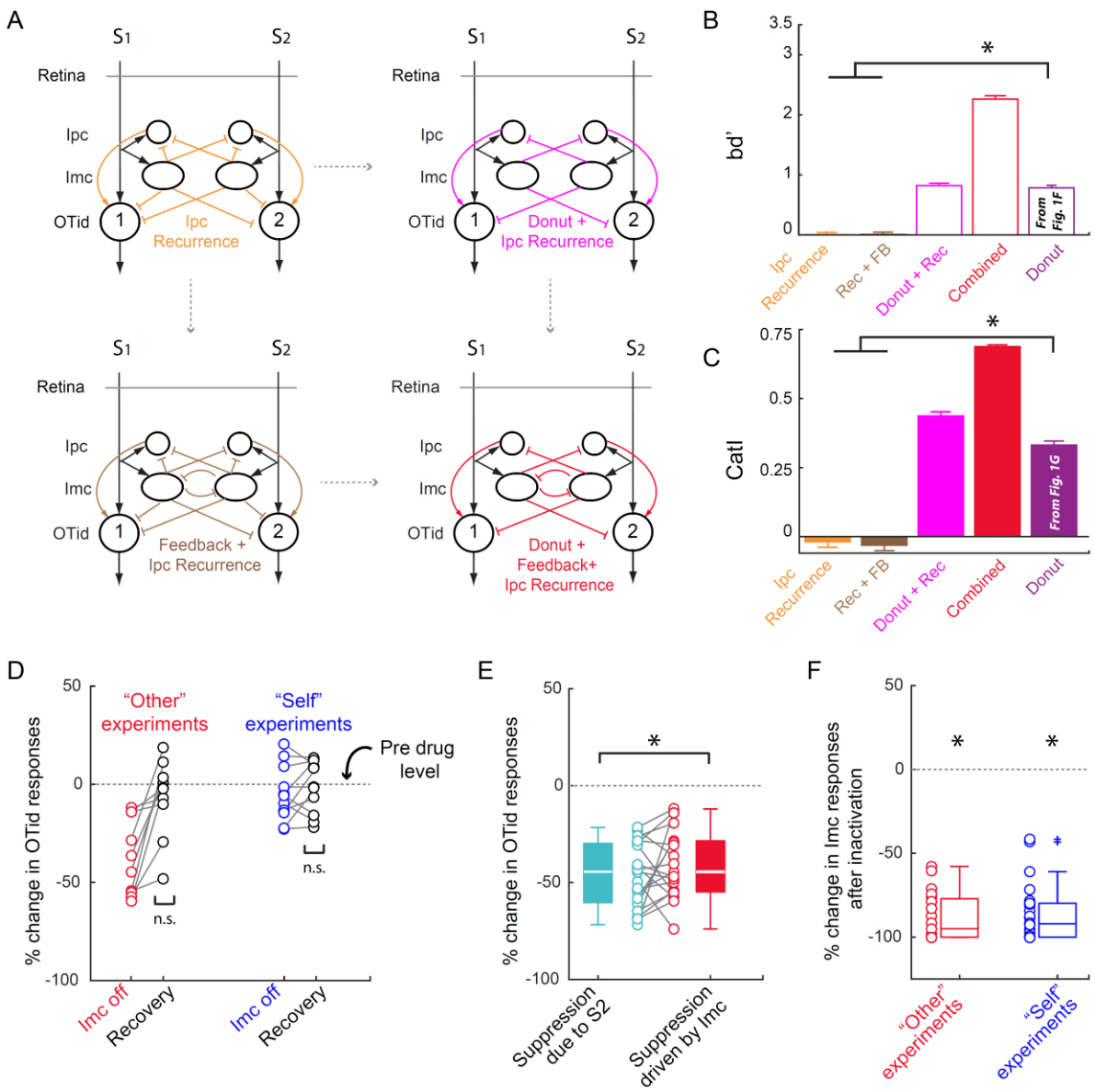


Fig. S2. Donut-like inhibition surpasses other circuit motifs even in the case of biologically grounded model (model based on barn owl Imc-Ipc-OT connectivity and function), and supporting data for Fig. 2 (experiments).

(A) *Baseline + recurrent excitation model, similar to the one in Fig. 1D, top-left, but with Ipc providing recurrent amplification, and receiving inhibition from Imc. Bottom-left: Adding feedback inhibition to model in top-left. Top-right: Adding donut-like motif to model in top-left. The donut-like motif is 'added' by setting "self" inhibition to zero; this is done both for the*

direct inhibitory pathway from Imc → OTid, and the indirect pathway: Imc → Ipc → OTid. Bottom-right: Adding both feedback inhibition and donut-like inhibition. ***(B)*** *bd' for these four models in A, computed on the response profiles of model neuron 1 obtained by simulating with the same two-stimulus morphing protocol used in Fig. 1E. Shown ghosted in for comparison are the results from the circuit model in Fig. 1C, top-right, that has just the donut-like motif without any recurrent amplification or feedback inhibition. '*': p = 9.5e-10, ANOVA followed by HBMC correction for multiple comparisons.* ***(C)*** *Same as B, but CatI. Results show that the supremacy of the donut-like motif over recurrent amplification alone, or the combination of that with feedback inhibition, in generating categorical response profiles still holds in the (biologically grounded) model of the barn owl midbrain circuit. '*': p = 9.6e-10, ANOVA followed by HBMC correction for multiple comparisons.* ***(D)*** *Recovery of OTid responses from kynurenic acid iontophoresis for experiments in Fig. 2. Data on left: "Other" experiments. OTid responses revealed significant change during iontophoresis (red dots; subset of the data reproduced from Fig. 2L), but returned to pre-drug baseline (horizontal line) in recovery (black dots; p=0.32, t-test against 0). Recovery data obtained 15 min after iontophoretic eject current was switched to retain current; Methods). Data show recovery, demonstrating that the effects reported Fig. 2 are due specifically to drug iontophoresis/Imc inactivation. Data on right: "Self" experiments. OTid responses showed no significant change during iontophoresis (blue dots; reproduced from Fig. 2L), and stayed around zero in recovery (black dots; p=0.58, t-test against 0.* ***(E)*** *"Other" experiment. Comparison of suppression provided by Imc with that due to S2 (i.e., the maximum amount of suppression experienced by the OTid neuron in this stimulus protocol). This is done by comparing (i) % suppression of OTid responses by stimulus S2 when Imc is intact, computed by comparing OTid responses to S1 alone versus the paired presentation of S1 and S2 (teal;(15)),*

with (ii) % suppression of OTid responses to paired presentation of S1 and S2 produced by Imc inactivation, computed by comparing OTid responses in the Imc-intact to Imc-off conditions (red; data reproduced from Fig. 2L). Consistent with (32), nearly all the suppression due to competitor S2 (teal) is supplied by Imc (red) (p= 0.22, t-test, teal vs. red), verifying that Imc supplies powerful "other" inhibition. ***(F)*** *Quantifying the effectiveness of Imc inactivation by kynurenic acid iontophoresis in the "other" (red) and "self"-inhibition (blue) experiments. In both cases, the inactivation was highly effective. Red: median % change in response; median = 95%, 95% CI = [87%, 103%], p = 3.8 e-6, sign test; Blue: median % change in responses median =0.92, 95% CI = [86%, 98%], p = 7.45e-9, sign test.*

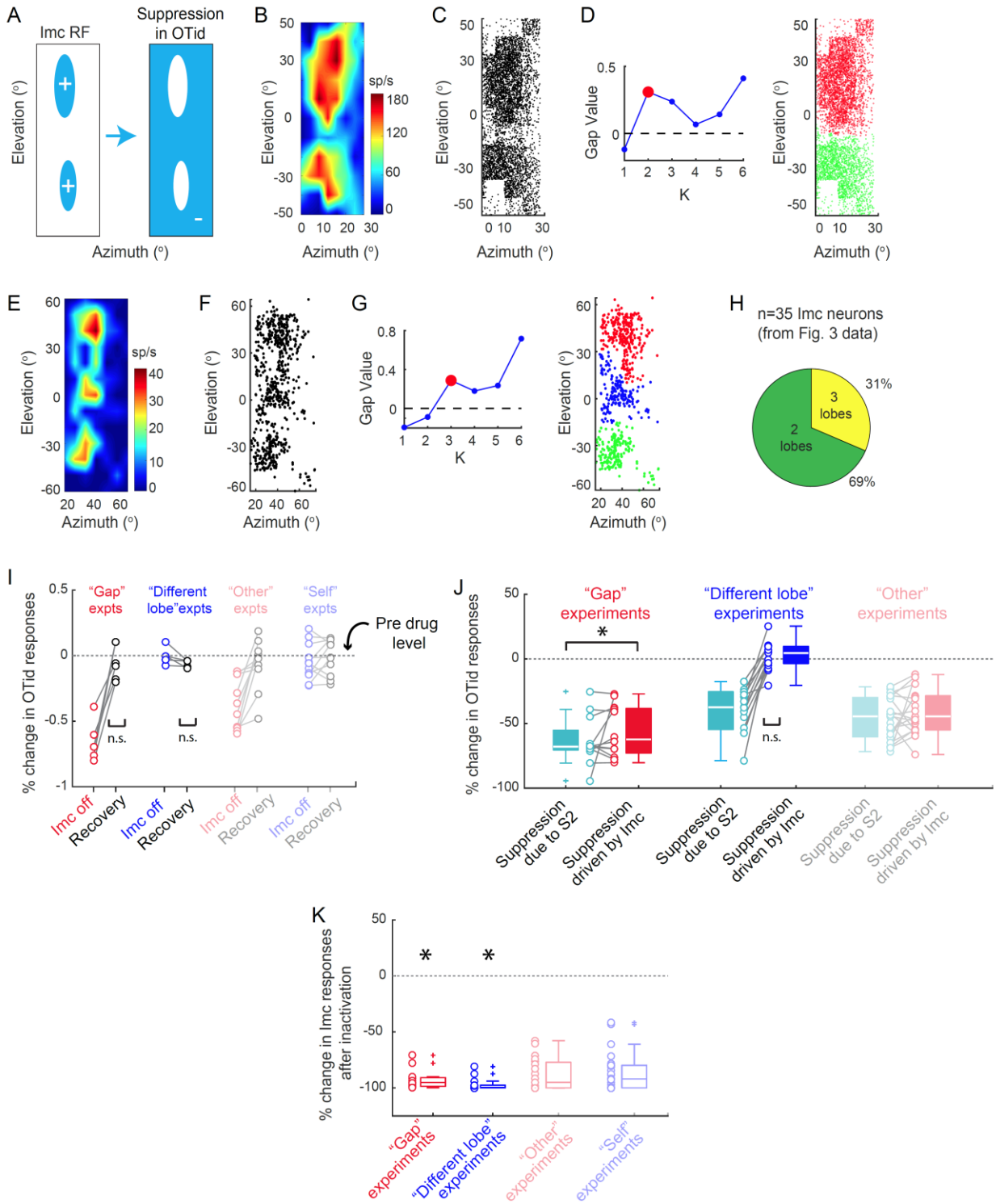


Fig. S3. Detecting the number of Imc RF lobes, and other supporting experimental data for Figure 3.

(A) *Schematic representation of a multi-holed donut. Left: Two-lobed RF of a putative Imc neuron (left; blue ovals are the RF lobes). Right: Pattern of inhibition created by this neuron in the OTid such that it spares all the locations form which it receives input. This complementary output projection pattern results in two "holes" – areas where no inhibition arrives due to the neuron on the left.* ***(B)*** *Imc RF reproduced from Fig. 3A-left.* ***(C)*** *Firing rate map from B converted to density of points in 2-D plane following published procedures (40).* ***(D)*** *Left: Density peaks clustering method (51) is applied to the data in B, forcing the method to yield either 1 cluster, 2 clusters, … upto 6 clusters. Following that, the gap statistic model selection metric (52) is applied to the clustering results to identify the optimal number of clusters in the data. The optimal number is the number for which the "gap value" exceeds 0 for the first time; here it is 2. Right: Same as B, but with the two colors indicating the two distinct lobes identified by the clustering method (+ gap statistic) as the optimal two best clusters in the RF data.* ***(E-G)*** *Same as B-D, but for Imc RF in Fig. 3G-left; determined to be three-lobed.* ***(H)*** *Pie-chart showing proportion of two-lobed vs. three-lobed RFs across all the multi-lobed Imc neurons recorded in this experiments (Fig. 3).* ***(I)*** *Recovery of OTid responses from kynurenic acid iontophoresis for experiments in Fig. 3. Conventions as in Fig. S2D; data from Fig. S2D reproduced here for comparison. Data show recovery ("Gap" expt: p =0.39, t-test of black data against 0; "Other" lobe expt: p=0.29, ranksum test of black data against 0), demonstrating that any effects reported Fig. 3 are due specifically to drug iontophoresis/Imc inactivation.* ***(J)*** *Comparison of suppression provided by Imc with that due to S2 (i.e., the maximum amount of suppression experienced by the OTid neuron in this stimulus protocol) in the "gap" (red data) and "different lobe" (blue data) experiments. Conventions as in Fig. S2E; data from Fig. S2E reproduced here for comparison. "Gap" expt: Consistent with (32), nearly all the suppression*

due to competitor S2 (teal) is supplied by Imc (red) (p= 0.47, teal vs. red, ranksum test).

<u>"Different lobe" expt</u>: The Imc neuron that is being inactivated does NOT provide any of the inhibition to the OTid neuron due to S2 (p = 0.35, blue dots against 0, t-test). This clearly demonstrates, consistent with the predictions of (40), that a different Imc neuron exists, which has one RF lobe encoding S2's location, but other RF lobes (if there are others for that neuron) not encoding S1's location, thereby delivering inhibition to that location! This is one of the 'signature' properties for the combinatorial encoding of space described in that study. ***(K)*** *Quantifying the effectiveness of Imc inactivation by kynurenic acid iontophoresis in the "gap" (red) and "different lobe" (blue) experiments. "Gap" expt: p =1.13 e-12 t-test against 0; "Other" lobe expt: p =1.53e-5, signtest. Conventions as in Fig. S2F; data from Fig. S2F reproduced here for comparison.*

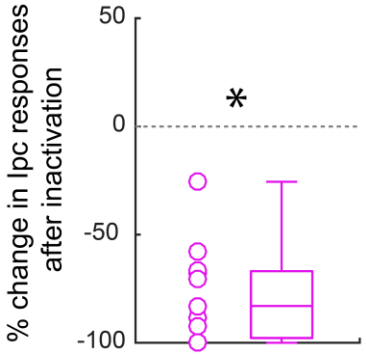


Fig. S4. Ipc is effectively inactivated (experiments re. Figure 4)

Quantifying the effectiveness of Ipc inactivation by kynurenic acid iontophoresis in the "necessity of donut-like inhibition" experiment. The inactivation was highly effective: (mean % change in responses = -77.42%, sd =22.82% p =5.3e-7, t-test against 0).