Neural circuit mechanism for generating categorical representations

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

*Corresponding and Lead author; shreesh.mysore@jhu.edu

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 far surpasses motifs of feedback inhibition, recurrent excitation, and divisive normalization invoked frequently 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. Indeed, the functional pattern of neural inhibition in the midbrain forms an exquisitely structured 'multi-holed' donut consistent with this network's combinatorial inhibitory function. Moreover, simulation results reveal a generalizable neural implementation of the donut-like motif for categorization across brain areas. Self-sparing inhibition may be a powerful circuit module central to categorical selection.

INTRODUCTION

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 (Bathellier et al., 2012; Freedman and Assad, 2006; Freedman et al., 2001; Hirokawa et al., 2019; Jiang et al., 2018; Jovanic et al., 2016; Mysore et al., 2011; Mysore and Knudsen, 2011a; Niessing and Friedrich, 2010; Xin et al., 2019). 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 overall robustness to sensory and neural noise, of the decoding of the winner from neural responses, can be quantified using boundary discriminability (*bd'*) - the discriminability between two competing options on either side of the selection boundary (d-prime; Methods): bd' is higher for more categorical response profiles (Fig. 1A and 1B-left, open symbols: red vs. orange). However, *bd'* is also higher for response profiles that are more robust-to-noise *without* being more categorical - for instance, scaled response profiles (Fig. 1A- orange vs blue). By contrast, a metric that is insensitive to scaling and that captures, specifically, the strength of categorization by response profiles is the commonly used categorization index ((Freedman and Assad, 2006; Mysore and Knudsen, 2011a); it operates effectively on the normalized response profiles; Fig. 1B-middle inset). A modified version, *CatI,* takes into account response variability

as well: it compares the average between-category discriminability with the average within-category discriminability (Fig. 1B-right, filled symbols; methods). Thus, both scaling-up of responses as well as enhancing their categorical nature can improve the overall robustness-to-noise of the neural signaling of the winner (Fig. 1AB, black vs. red). Biologically, however, the limited dynamic ranges of neural responses (which, in turn, limit the range of scaling possible), as well as the inherent discreteness of categorical responses, make categorical neural representations a preferable solution to promote robust-to-noise selection.

Despite the clear utility of categorical representations, their pervasiveness in brain areas (Bathellier et al., 2012; Freedman and Assad, 2006; Freedman et al., 2001; Hirokawa et al., 2019; Jiang et al., 2018; Jovanic et al., 2016; Mysore et al., 2011; Mysore and Knudsen, 2011a; Niessing and Friedrich, 2010; Xin et al., 2019), as well as the range of theoretical formulations and computational models proposed to account for them (Engel et al., 2015; Machens et al., 2005; Mysore and Kothari; Wang, 2012), identifiable 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.

RESULTS

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 (Bollimunta and Ditterich, 2012; Hartline et al., 1956; Mysore et al., 2010; Olsen et al., 2010), 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

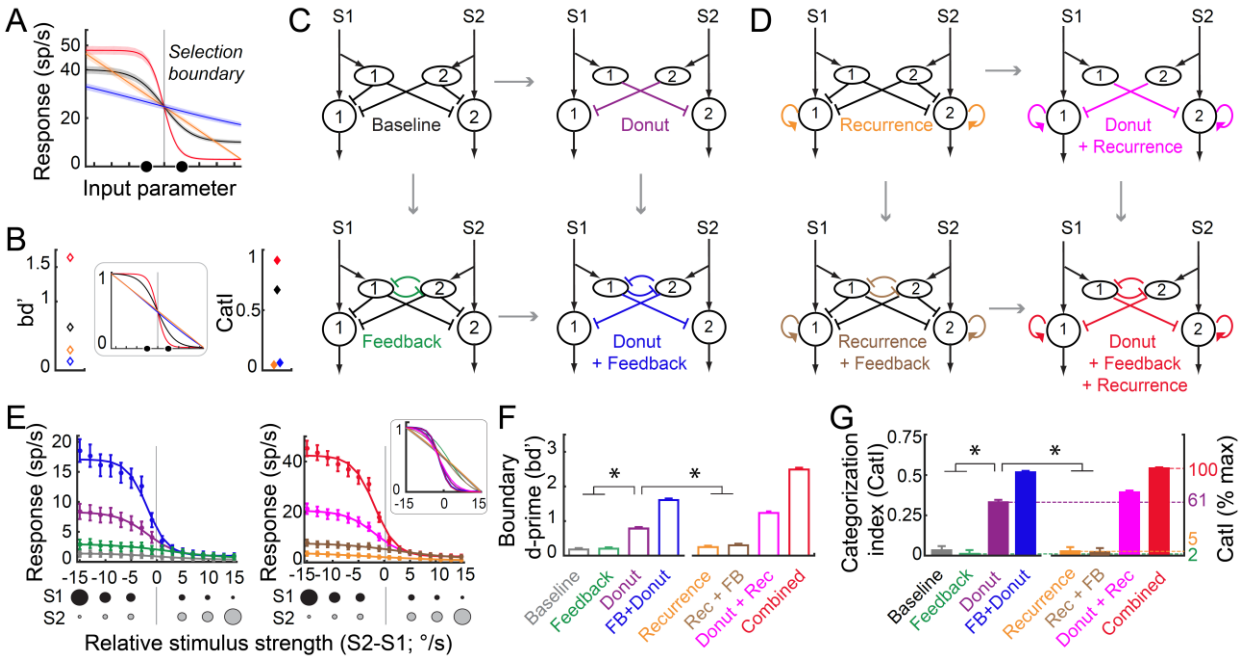
options, may play a key role. This was motivated by extensive modeling of categorical decision-making (Machens et al., 2005; Wang, 2008), work on direction selectivity in the retina (Chen et al., 2016), as well as work on spatial selection in barn owls (Mysore and Knudsen, 2012). The second hypothesis 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 (Sereno and Ulinski, 1987) 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;(Bollimunta and Ditterich, 2012; Hartline et al., 1956; Olsen et al., 2010)), 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; (Freedman and Assad, 2006; Mysore and Knudsen, 2011a; Niessing and Friedrich, 2010)). 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,

87 we compared both scaling-dependent robustness-to-noise (*bd'*) as well as the scaling-insensitive

88 categorization strength (*CatI)* across the response profiles (ANOVA with HBMC correction;

89 Methods).



90

91 ***Fig. 1. Modeling: Donut-like inhibition surpasses other hypothesized circuit motifs in***

92 ***generating categorical responses to competing stimuli.***

93 ***(A)*** *Schematic showing 4 different mathematically generated response profiles, as a function of*

94 *continuously varying input. Vertical line: categorization or selection boundary. Translucent*

95 *band: variability in responses; fano factor of 6 used to generate these responses (Methods).*

96 *Black filled dots: Two competing inputs just straddling the selection boundary: nearly equal in*

97 *input value (at a distance of 3 units from the selection boundary), but belonging to different*

98 *categories.* ***(B)*** *<u>Left</u> (open symbols): Boundary discriminability (bd'); defined as d-prime*

99 *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; accounts for response variability as well; Methods) for response profiles in A. (Freedman and Assad, 2006; Mysore and Knudsen, 2011a). ***(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; (Albantakis and Deco, 2009; Machens et al., 2005; Wang, 2002)).

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

144 donut-like motif, they were nearly ineffective, either individually or together, at signaling the
145 winning option categorically in the presence of input ambiguity and neural noise (Fig. 1FG:
146 green, orange, brown).

147 **Functionally donut-like inhibitory motif operates in the barn owl midbrain**

148 We wondered, next, if the donut-like inhibitory motif operates within a neural selection circuit.
149 An excellent site in the brain to investigate this is the midbrain spatial selection network in
150 vertebrates (Knudsen, 2011; Mysore and Knudsen, 2011b). It includes the superior colliculus, SC
151 (optic tectum, OT, in non-mammals), a sensorimotor hub that is required for target selection for
152 spatial attention (Lovejoy and Krauzlis, 2010; McPeek and Keller, 2004). Neurons in the
153 intermediate and deep layers of the SC (SCid/OTid) encode a topographic spatial map of
154 stimulus priority (priority = physical salience + behavioral relevance) (Fecteau and Munoz,
155 2006), and OTid neurons signal the highest priority stimulus categorically*:* they respond with a
156 high firing rate when the stimulus inside their RF is the one with highest priority, but their
157 responses are suppressed to a low level when the RF stimulus is no longer the one with highest
158 priority (Mysore et al., 2011; Mysore and Knudsen, 2011a, 2014). The source of this competitive
159 response suppression in the OTid is a group of GABAergic midbrain tegmental neurons called
160 Imc in birds (isthmi pars magnocellularis; (Wang et al., 2004)): Imc inactivation abolishes all
161 competitive interactions in the OTid and thereby its ability to signal the winner (Marin et al.,
162 2007; Mysore and Knudsen, 2013).

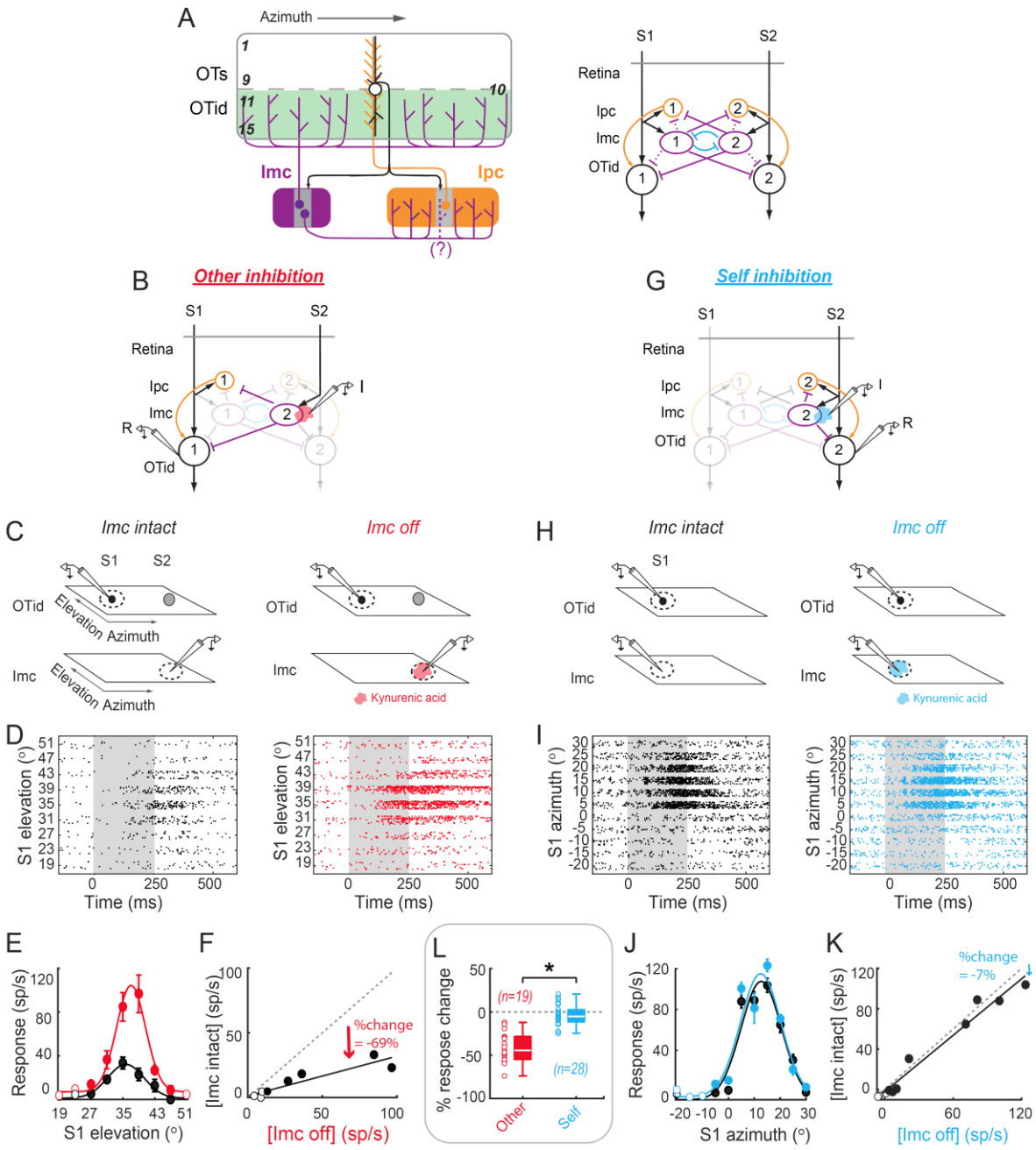


Fig. 2. Barn owl midbrain selection network contains functional donut-like inhibitory motif operating along azimuthal space.

(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*

168 *OT tissue, from left to right, encode individual locations in space (here, azimuth)*
169 *topographically (Knudsen, 1982). Imc (purple): Isthmi pars magnocellularis, inhibitory neurons*
170 *in midbrain tegmentum. Ipc (orange): Isthmi pars parvocellularis, cholinergic neurons. Imc and*
171 *Ipc receive focal input from OT10 (dark grey projections). Ipc projects back focally to OT*
172 *(orange projections). Imc projects back broadly across OTid and Ipc (purple projections);*
173 *thought to spare the portion of OT providing input (absence of purple projections in the column*
174 *occupied by black OT10 neuron; '?' (Wang et al., 2004): unknown if congruent portion of Ipc*
175 *space map is also spared of Imc projections. Right: Network model showing the OT-Imc-Ipc*
176 *circuit; conventions as in Fig. 1. Simulations from this model in Fig. S2. Dashed lines: Unknown*
177 *if these connections exist functionally.* ***(B-F)*** *Measurement of the strength of net "other"*
178 *inhibition from Imc → OTid with paired recordings in Imc and OTid. "Net" indicates the*
179 *combined inhibition due to both the direct (Imc → OTid) and indirect (Imc → Ipc → OTid)*
180 *pathways (text). "Other" indicates that the OTid neuron encodes for (distant) spatial locations*
181 *outside Imc neuron's RF. (B) Experimental setup. Iontophoresis and recording electrode (I) in*
182 *the portion of Imc (encoding for stimulus S2); recording electrode (R) in the portion of OTid*
183 *encoding for distant location (and stimulus S1). (C) Schematic of OTid and Imc space maps*
184 *(quadrilaterals) showing RFs of neurons being recorded (dotted ovals) and stimulus protocol*
185 *(black filled dot – S1; gray filled dot – S2). S1 and S2 are looming visual stimuli of fixed contrast*
186 *but different loom speeds (strength; Mysore 2010); S1 = 9.6 °/s, S2 = 19.2°/s (Methods). (D)*
187 *Raster responses of example OTid neuron to paired stimulus protocol in C, in the Imc-intact*
188 *condition (left column; black data) and Imc-off condition (right column – red data). Imc*
189 *inactivation by (reversible) iontophoresis of kynurenic acid (a pan-glutamate receptor*

blocker;(Mysore and Knudsen, 2013); 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; (Goddard et al., 2014; Wang et al., 2004)), and the latter, as response amplification by a group of

211 cholinergic neurons, Ipc (Fig. 2A; orange), that connect in a point-to-point manner with the OT
212 (Asadollahi and Knudsen, 2016; Marin et al., 2005; Wang et al., 2006).

213 What, however, of the donut-like motif? Imc neurons are known to receive input from a
214 focal portion of the OT (within layer 10; OT_{10}), but to suppress OTid activity broadly across the
215 space map by sending projections via two pathways (Fig. 2A): (a) directly to the OTid (layers
216 11-15) (Fig. 2A – purple;(Wang et al., 2004)), and (b) indirectly to the OTid by inhibiting the
217 potent point-to-point amplifiers of OTid activity, namely the cholinergic Ipc (Fig. 2A –orange;
218 (Marin et al., 2007; Wang et al., 2004)). Incidentally, simulations with a circuit model in which
219 the recurrent amplifier is also the recipient of powerful inhibition, as in the OT-Imc-Ipc circuit,
220 confirm the superiority of the donut-like motif for producing categorical representations (Fig.
221 S2A-C).

222 Whereas anatomical tracing studies (Wang et al., 2004) have indicated that the direct
223 projections from the Imc to the OT may spare the portion of the OT providing input to the Imc
224 (Fig. 2A; absence of purple projections in portion of OT around black neuron), this not been
225 established functionally. Crucially, whether or not the indirect inhibitory pathway involving the
226 Ipc exhibits the donut-like motif is unknown (Fig. 2A, purple '?'). This is critical because this
227 indirect pathway is known to constitute the dominant route of inhibition from Imc to OTid
228 (Wang et al., 2004): the majority of Imc projections target the Ipc (rather than the OTid; (Wang
229 et al., 2004)), and Ipc provides substantial amplification of OTid activity (1.43x, on average;
230 (Asadollahi and Knudsen, 2016)).

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

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

237 To measure the strength of "other" inhibition, we recorded the responses of OTid neurons
238 to a stimulus (S1) inside their receptive field (RF; Fig. 2B, C-left; Methods) while
239 simultaneously presenting a competing stimulus outside the RF (S2; at a distant azimuthal
240 location from S1; Methods). Responses to S1 in the OTid are known to be suppressed by a
241 distant S2, by an amount depending on their relative strength (Mysore et al., 2010; Rizzolatti et
242 al., 1974). Notably, this suppression is known to be abolished upon focally inactivating the
243 portion of Imc representing S2, revealing Imc as the primary source of competitive suppression
244 in the OTid (Marin et al., 2007; Mysore and Knudsen, 2013). Therefore, to estimate the strength
245 of inhibition delivered by Imc neurons to OTid neurons encoding 'other' locations, we compared
246 the responses to paired presentation of S1 and S2 when Imc was intact (Fig. 2C-left) versus when
247 the portion of Imc encoding S2 (spatially mismatched with the OTid recording site), was focally
248 and reversibly inactivated (Fig. 2C-right). Inactivation was achieved by the blockade of
249 excitatory synaptic inputs onto Imc neurons through iontophoresis of kynurenic acid (Methods).
250 The difference in OTid responses between the Imc-intact (Fig. 2D-left; 2E-black) and the Imc-
251 inactivated (Fig. 2D-right; 2E-red) conditions quantified the amount of suppression due to Imc
252 onto the "other" OTid location, thereby directly estimating the strength of net "other" inhibition.
253 (No change in OTid responses would indicate zero competitive suppression by Imc onto that
254 OTid neuron.)

255 We found that Imc neurons exerted strong inhibition onto OTid neurons encoding for
256 distant, non-matched spatial locations (Fig. 2F, L-red; mean strength = - 40.47 % +/- 17.70 %,
257 n==19 paired neurons; p=9.43e-9, t-test against 0; mean distance between centers = 26.74 °). We
258 verified that the results were specifically due to Imc inactivation by observing that OTid
259 responses to paired S1 and S2 after recovery from iontophoresis (measured 15 min after the drug
260 was turned off; Methods) returned to pre-drug levels (Fig. S2D). Indeed, the suppression
261 provided by Imc accounted for nearly all the suppression exerted by the stimulus S2 (Fig. S2E).
262 In these experiments, Imc was inactivated effectively (median = 95%, 95% CI = [87%, 103%]; p
263 = 3.8e-6, sign test, n=19; Fig. S2F).
264 To measure the strength of self-inhibition, we recorded the responses of OTid neurons to
265 a single stimulus (S1) presented inside their RF (Fig. 2G; Methods). We compared these OTid
266 responses when Imc was intact (Fig. 2H-left) versus when the portion of Imc encoding S1, i.e.,
267 spatially matched to the OTid recording site, was focally and reversibly inactivated (Fig. 2H-
268 right). Any difference in OTid responses between the Imc-intact (Fig. 2I-left; 2J-black) and the
269 Imc-inactivated (Fig. 2I-right; 2J-blue) conditions quantified the amount of suppression due to
270 Imc onto the "self" OTid location, thereby directly estimating the strength of net 'self-
271 inhibition'.
272 We found that Imc neurons exerted no significant inhibition onto OTid neurons encoding
273 for overlapping spatial locations (Fig. 2K, L-blue; mean strength = -3.7 %, s.d. = 12.2%, n = 28
274 neuron pairs; p = 0.12, t-test against 0; mean distance between centers = 2.86 °). We verified that
275 these results were not due to ineffectiveness of iontophoresis by observing that the suppression
276 of Imc responses by kynurenic acid was substantial (Fig. S2F), and not distinguishable from that

277 in the “other” case (Fig. S2F; p = 0.68, ranksum test, Imc suppression by drug in “self” vs.
278 “other” cases). Therefore, strength of net “self” inhibition in the OTid was substantially weaker
279 than the strength of net competitive inhibition (Fig. 2L, red vs. blue; p = 7.8e-11, two sample t
280 test with HBMC correction; Methods). Together, these findings demonstrated the presence of a
281 (net) donut-like, spatial pattern of inhibition across the OTid azimuthal space map (i.e. when S1
282 and S2 were separated along the azimuth).
283

284 **Donut-like inhibitory motif in the barn owl midbrain is multi-holed**

285 Imc neurons are strikingly asymmetric in their encoding of elevational versus azimuthal space.
286 The majority (67%) exhibit RFs with multiple discrete firing fields (“lobes”) distributed along
287 the elevation, but not the azimuth (multilobed RFs exhibit upto 3 lobes along the elevation;
288 (Mahajan and Mysore, 2018); Fig. 3A,G-left panels).This unusual RF structure has been shown
289 to be essential for Imc to achieve selection at all possible pairs of spatial locations in the face of
290 scarcity of its neurons, and it does so does using a novel, combinatorially optimized inhibitory
291 strategy (Mahajan and Mysore, 2018).

292 A direct consequence of multilobed encoding of elevational space is that there are gaps
293 between the lobes of an Imc RF, constituting locations that are outside that neuron’s RF (Fig.
294 3AB-left panels; light red bar). If the donut-like inhibitory motif is to operate also for
295 (categorical) selection among competing stimuli separated along the elevation, then the spatial
296 pattern of inhibition must respect a very strict condition: namely, that, multilobe Imc neurons
297 ought to send strong competitive inhibition to OTid neurons encoding locations in the gaps
298 between RF lobes, but weak or no inhibition to OTid neurons encoding locations within the other

299 RF lobes of the same Imc neurons (weak self-inhibition), predicting a 'multi-holed' donut (Fig.
300 S3A).

301 To test experimentally if this strict requirement holds true, we again made dual
302 extracellular recordings in the OTid and Imc. We first recorded from an Imc neuron, mapped out
303 its spatial RF, and applied previously published analyses to determine if it was a multilobed RF
304 (two-lobed RF in Fig. 3AB – left panels and Fig. S3B-D; three-lobed RF in Fig. 3GH – left
305 panels and Fig. S3E-G; Methods;(Mahajan and Mysore, 2018)). Next, we positioned a second
306 electrode in the OTid such that the spatial RF of the OTid neuron was either centered within the
307 gap between Imc RF lobes (Fig. 3AB – right panels and Fig. 3C-left, light red bar) or it
308 overlapped one of the lobes of the Imc neuron's RF (Fig. 3GH – right panels and Fig. 3I – left,
309 light blue bar). We, then, recorded the responses of the OTid neuron to a stimulus inside its RF
310 (S1; Fig. 3C-left and I-left; Methods) while simultaneously presenting a competing stimulus (S2)
311 at a distant location along the elevational axis such that S2 was within a (different) lobe of the
312 RF of the Imc (Fig. 3C-left – S1 in the Imc RF gap – light red bar and S2 within an Imc RF lobe;
313 Fig. 3I-left – S1 within one Imc RF lobe and S2 within a different one – light blue bar). We
314 compared OTid responses when Imc was intact (Fig. 3CI-left panels) versus when the portion of
315 Imc encoding S2 was focally and reversibly inactivated (Fig. 3CI-right panels). As before, any
316 observed response differences directly estimated, respectively, the strength of net inhibition
317 exerted by the Imc neuron onto the OTid neuron encoding the gap location ("gap" inhibition), or
318 the strength of net inhibition exerted by the Imc neuron due to a stimulus within one of its lobes,
319 onto the OTid neuron encoding for a location within a different lobe of the Imc's RF ("different
320 lobe" inhibition).

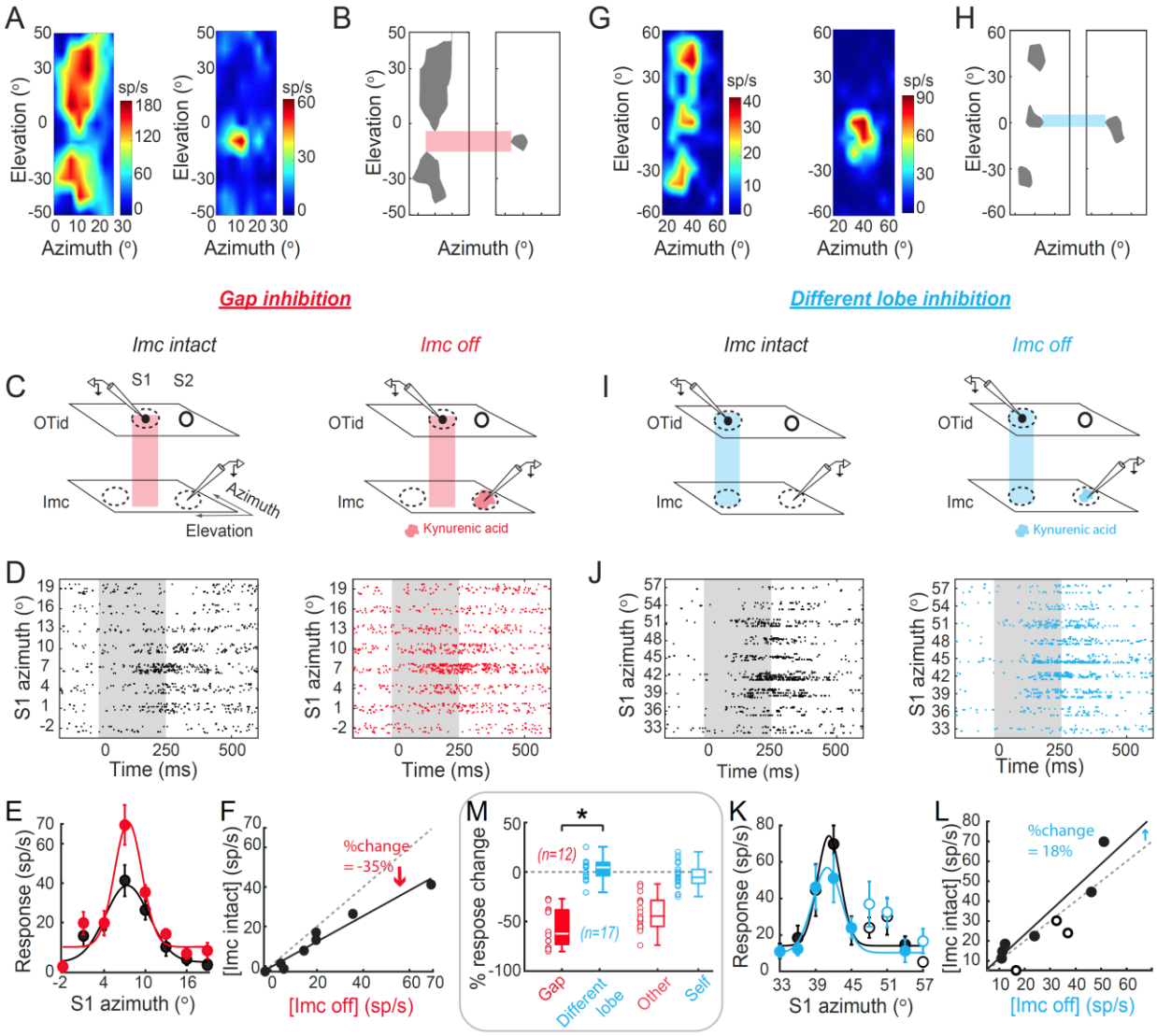


Fig. 3. Midbrain selection network in barn owl contains multi-holed donut-like inhibitory motif operating across 2-D sensory space (azimuth x elevation).

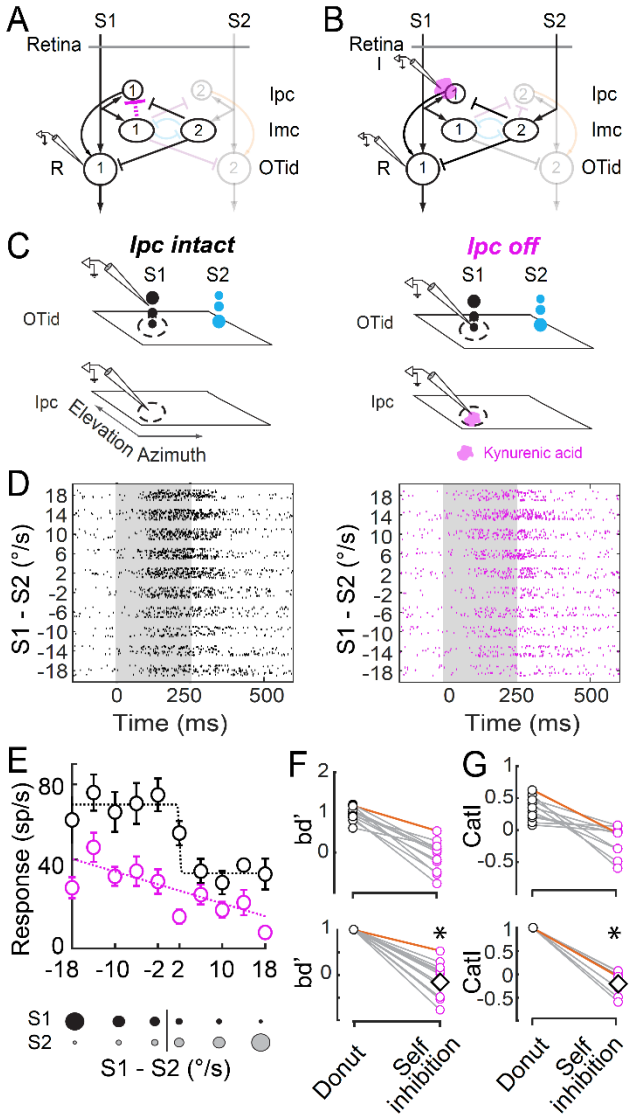
(A-F) *Measurement of the strength of net "gap" inhibition from Imc → OTid with paired recordings in Imc and OTid. (A) Spatial receptive fields (RFs) of an Imc neuron (left) and an OTid neuron (right) from a paired Imc-OTid recording experiment. Imc RF is two-lobed (text; Fig. S3B-D; Methods; (Mahajan and Mysore, 2018)). 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*

329 *horizontal bar: highlights the relative position of OTid RF to Imc RF lobes. (C-F) Conventions*
330 *as in Fig. 2C-F. C, Red vertical bar indicates that OTid RF is in the gap between Imc RF lobes.*
331 *F, Line: Best fit straight line; slope = 0.65, R^2 =0.95. % change in responses directly estimates*
332 *the strength of inhibition at this OTid neuron due to Imc (here, 35%).* ***(G-L)*** *Measurement of the*
333 *strength of net "different lobe" inhibition from Imc → OTid with paired recordings in Imc and*
334 *OTid. Conventions as in A-F. (G, H) Three-lobed spatial RF of an Imc neuron (left; Fig. S3E-G;*
335 *Methods). RF of OTid neuron overlaps with one of the lobes of Imc neuron's RF (blue bar). (L)*
336 *Line: Best fit straight line; slope = 1.18, R^2= 0.9. % change in responses directly estimates the*
337 *strength of inhibition at this OTid neuron due to Imc (here, -18%).* ***(M)*** *Population summary of*
338 *strength of net "gap" inhibition (red; n=12), and strength of net "different lobe" inhibition*
339 *(blue; n=17) from Imc → OTid. Open red and blue box plots: reproduced from Fig. 1L. p<1e-5*
340 *(red vs. 0), p=0.35 (blue vs. 0); p=<1e-5 (red vs blue), paired t-tests with HBMC correction.*
341 *See also Fig. S3.*
342
343 We found that Imc neurons exerted strong competitive inhibition onto OTid neurons that
344 encoded locations in the gaps between the Imc RF lobes (Fig. 3D-F, M-red), but weak
345 competitive inhibition onto OTid neurons that encoded for locations within a different lobe of the
346 Imc's RF (Fig. 3J-L, M-blue). The mean strength of "gap" inhibition (along elevation) was -
347 55.88% (Fig. 3M-red, s.d. = 20%, n==12 neuron pairs; p = 1.02 e-6, t-test against 0 with HBMC
348 correction; Fig. S3IJ-red). The mean strength of "different lobe" inhibition was 2.43 % (Fig. 3M-
349 blue, s.d. = 10.3 %, n==17 neuron pairs; p = 0.35, t-test against 0 with HBMC correction; Fig.
350 S3IJ-blue), and was not distinguishable from the average "self" inhibition (Fig. 3M- blue vs.

351 translucent blue; p-value = 0.09; two sample test with HMBC correction). Imc was inactivated
352 effectively in both sets of experiments (median = 98.2% median, 95% CI = [96.46% 99.99%];
353 Fig. S3K).
354 Taken together, these results demonstrated that the Imc implements precisely organized,
355 multi-holed donut-like patterns of inhibition onto the OTid space map, with each Imc neuron's
356 net inhibitory effect creating a pattern complementary to its (multilobed) RF structure (Fig.
357 S3A).
358
359 **Donut-like inhibitory motif is required for categorization by barn owl midbrain**
360 Considering the wiring complexity of the (multi-holed) donut-like connectivity between Imc
361 (and Ipc) and OT, we wondered if it served a functional purpose. Specifically, we investigated
362 whether it was necessary for the categorical signaling of the highest priority stimulus by the
363 OTid. To test this, we reasoned that causally introducing "self" inhibition onto an OTid neuron
364 to disrupt the donut-like pattern of inhibition would be most effective if achieved via the indirect,
365 but potent, pathway, i.e., by increasing "self" inhibition onto Ipc, (Fig. 4A). As the net "self"
366 inhibition at most OTid neurons is weak or zero (Figs. 2L and 3M), activating Imc → Ipc
367 projections (or Imc → OTid projections) for spatially aligned neuron pairs would be infeasible
368 because such projections are unlikely to exist. Instead, we mimicked this effect by focally
369 inactivating Ipc neurons (Fig. 4B-right), while recording simultaneously the responses of
370 spatially matched OTid neurons to the two-stimulus morphing protocol used to study categorical
371 signaling in the OTid (Fig. 4C left vs. right; protocol shown in Fig. 4E–bottom, same as Fig. 1E–
372 bottom; Methods;(Mysore and Knudsen, 2011a)).

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,*



395 *magenta: Ipc-off condition. Dashed lines: best fitting sigmoid or straight line to data, chosen*
396 *based on AIC criterion. Black: AIC (sigmoid) = 29.07, AIC (line) = 40.92; magenta: AIC*
397 *(sigmoid) = 37.75, AIC (line) = 34.67.* ***(F, G)*** *Population summary of effect of Ipc inactivation*
398 *on bd' (F) and CatI (G); n=11 neuron pairs; data in orange are from example neuron pair in D-*
399 *E. Top panels: measured values. Bottom panels: Data in top panels replotted after normalizing*
400 *to Ipc-intact values. Diamond: average of magenta data; p <1e-4 (bd'), t-test, and p <1e-4*
401 *(CatI), ranksum tests against 1 with HBMC correction in F and G. See also Fig. S4.*
402
403 We found that disrupting the donut-like inhibitory motif (by introducing "self" inhibition
404 onto the Ipc) caused a substantial reduction in categorical signaling in the OTid (Fig. 4D-E:
405 example neuron pair)**.** Across the population of tested neuron pairs, categorization as well as
406 robust-to-noise signaling were nearly abolished, with the median reduction in *CatI* of 104.7%
407 and a mean reduction in *bd'* of 112.42% (Fig. 4FG; n=11 neuron pairs; *bd'*: s.d. = 37.8%; p =
408 3.93e-9, t-test against 1 with HBMC correction; *CatI*: 95% CI of median = [91.52%, 117.92%];
409 p = 2.55 e-5, ranksum test against 1 with HBMC correction). Thus, the midbrain spatial selection
410 network not only contains the specialized donut-like inhibitory circuit motif, but also critically
411 depends on it for categorical, robust-to-noise signaling of the highest priority stimulus.
412
413 **Divisive normalization model is not effective for generating categorical representations**
414 The structured (multi-holed) donut-like organization of inhibition stands in contrast to another
415 computational mechanism that has been invoked in the decision-making literature, namely
416 divisive normalization (Albantakis and Deco, 2009; Carandini et al., 1997; Kouh and Poggio,

417 2008; Louie et al., 2013). It involves the pooling of inhibition due to various active channels,
418 with the inhibitory elements inhibiting each other, and with the pooled inhibition delivered
419 *uniformly* (rather than in a donut-like manner) to the output elements (Fig. 5A;(Carandini et al.,
420 1997; Kouh and Poggio, 2008)).

421 Our results (Figs. 2-3) show that the midbrain spatial selection network (OT-Imc-Ipc)
422 does not implement pooled divisive normalization for selection across space. To test, more
423 generally, whether the divisive normalization model is effective for generating categorical
424 response profiles, we simulated a circuit model of normalization (Fig. 5A) and obtained model
425 neuron responses to the two-stimulus morphing protocol (Fig. 5A; same protocol as in Figs. 1E
426 and 4). *CatI* was substantially lower for this model than that obtained with just the donut-like
427 inhibitory motif (Fig. 5C right panel: mean *CatI* = -0.05 for normalization model compared to
428 0.331 for donut model in Figure 1, p = 1.5 e-29, t-test with HMBC correction, gold vs. purple;
429 left panel: p = 1e-12, t-test of gold vs. purple,). Thus, the normalization mechanism is not
430 effective for generating categorical responses (consistent with findings from modeling in visual
431 cortex (Busse et al., 2009)). Indeed, it was the presence of self-inhibition, specifically, that
432 caused this circuit to be ineffective: a modified version of the circuit in Figure 5A which did not
433 include self-inhibition (Fig. 5B), was very effective, further attesting to the primacy of the donut-
434 like motif for categorization (Fig. 5C right panel: blue vs. gold data, mean *CatI* = 0.279 for
435 normalization model with donut compared to 0.022 for normalization model, p = 1.15e-14, t-test
436 with HBMC correction; left panel: p = 8.8e-19, t-test of blue vs. gold with HBMC correction).
437 (Incidentally, the donut-like motif is capable also of keeping the responses of the circuit in
438 check, in a sense, 'normalizing' the responses, as the overall drive increases (Fig. 1E)).

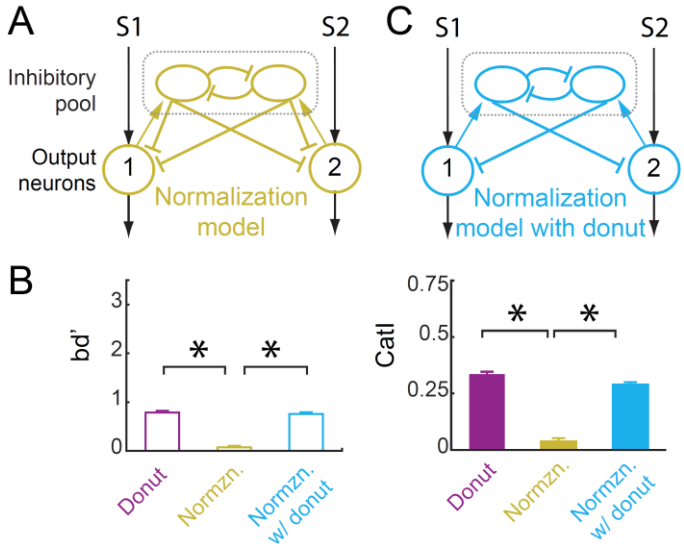


Fig. 5. Modeling: Divisive normalization model operating across the space map is not effective for producing categorical responses.

(A) *Schematic of normalization circuit model with pooled inhibition and inhibitory feedback (Carandini et al., 1997; Kouh and Poggio, 2008).* ***(B)*** *Schematic of normalization circuit sans self-inhibition.* ***(C)*** *Plot of bd' (left panel) and CatI (right panel) computed from the responses of neuron 1 in the normalization model (A; gold data) and the model sans self-inhibition (B; blue data), 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. Right panel (CatI): p = 5.2e-23 (gold vs. purple), p=8.8e-19 (gold vs. blue), paired t-tests with HBMC correction; left panel (bd'): p = 1.9e-31 (gold vs. purple), p=5.13e-31 (gold vs. blue), paired t-tests with HBMC correction.*

454 **Generalizable implementation of the donut-like motif**

455 Both in the barn owl brain and in the various models considered in Figures 1-4, the circuit
456 architectures included feedforward inhibition from inhibitory neurons to the 'output' neurons,
457 with the donut-like motif instantiated as the absence of feedforward self-inhibition. Several
458 models of decision making in humans and monkeys do include a feedforward inhibitory
459 component (Churchland and Ditterich, 2012; Ratcliff and McKoon, 2008; Roe et al., 2001;
460 Usher and McClelland, 2001). Examining those models in light of our findings reveals that they
461 indeed implement a donut-like motif, although this aspect of the circuit is not highlighted in the
462 models.

463 However, another established class of models of decision making does not include
464 feedforward inhibition, but only involves inhibition in a recurrent path between the competing
465 options (Machens et al., 2005; Wong and Wang, 2006). To test if feedforward implementation of
466 the donut-like motif was necessary for categorization or whether its efficacy generalized to
467 recurrent implementations as well, we simulated a version of our circuit model that did not
468 include feedforward inhibition, and that implemented the donut-like motif only via a recurrent
469 route (Fig. 6AB). For completeness, we simulated this model both without (Fig. 6A) and with
470 amplification (Fig. 6B). We found that this implementation of the donut-like motif also
471 successfully produced categorical representations (Fig. 6C; CatI=0.29 (black), 0.34 (light blue)).
472 Nonetheless, the feedforward implementation of the donut-like motif (Fig. 1CD bottom right
473 panels) offered additional benefits to categorization (Fig. 6C, right panel: black (Fig. 6A) vs.
474 dark blue (Fig. 1C-bottom–right), p = 1.7e-25; light-blue (Fig. 6B) vs. orange (Fig. 1D-bottom–
475 right), p = 2.5e-31; paired t-tests with HBMC correction), indicating that if a neural circuit (for

instance, the avian midbrain selection network) is able to implement this version of the donut, it can do better than one with a purely recurrent implementation.

These results, together with the (seemingly hidden) presence of the donut-like motif in the leading classes of decision-making models that have been used to account for categorical neural responses across brain areas and animal species (Bathellier et al., 2012; Churchland and Ditterich, 2012; Freedman and Assad, 2006; Freedman et al., 2001; Jiang et al., 2018; Machens et al., 2005; Niessing and Friedrich, 2010; Ratcliff and McKoon, 2008; Roe et al., 2001; Usher and McClelland, 2001; Wang, 2008; Wong and Wang, 2006), directly support the generality of the donut-like motif for producing categorical representations.

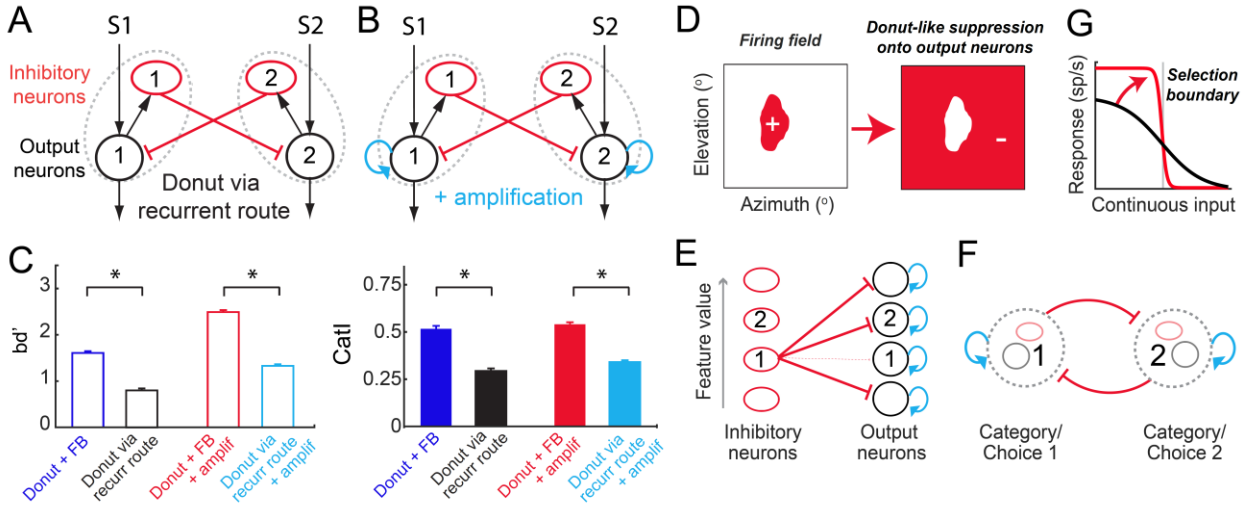


***Fig. 6.* Generalizable implementation of the donut-like motif.**

(A-B) *Circuit with implementation of donut-like motif purely via a recurrent route, i.e., in the absence of any feedforward inhibition without (A) and with amplification (B; blue arrow); compare with, respectively, Fig. 1CD bottom-right panels. Red ovals: inhibitory neurons, black circles: excitatory/output neurons, dashed gray ovals: populations of neurons representing each*

492 *stimulus or category.* ***(C)*** *Plot of bd' (left panel; black and light blue data) and CatI (right*
493 *panel; black and light blue data) computed from the responses of neuron 1 in circuits in A and B*
494 *to the standard two-stimulus morphing protocol (as in Fig. 1E). For comparison, the bd' and*
495 *CatI values for donut-like motif, from Fig. 1CD bottom-right panels, are reproduced here (dark*
496 *blue and orange data, respectively). '*': p<0.05. Left panel: black (A) vs. dark blue (Fig. 1C-*
497 *bottom–right), p = 1.7e-23; light-blue (B) vs. orange (Fig. 1D-bottom–right), p = 4.7e-34;*
498 *paired t-tests with HBMC correction. Right panel: black (A) vs. dark blue (Fig. 1C-bottom–*
499 *right), p = 1.7e-25; light-blue (B) vs. orange (Fig. 1D-bottom–right), p = 2.5e-31; paired t-tests*
500 *with HBMC correction.* ***(D-G)*** *Graphical summary. (D) Donut-like inhibition, i.e., inhibition (-)*
501 *driven by preferred inputs (+) and delivered to all non-preferred inputs, can be implemented in*
502 *neural circuits either in a feedforward manner (E) or in a recurrent manner (F), to generate*
503 *categorical representations (G). (E, F) Red ovals: inhibitory neurons, black circles:*
504 *excitatory/output neurons. E, dashed thin red arrow indicates absence of inhibitory projection. F*
505 *based on Figures 1-4; G is a simplified representation of model in B, in which populations of*
506 *neurons representing each category or choice (dashed grey circle) mutually inhibiting one*
507 *another.*
508

509 **Discussion**

510 Our results discover the power of a donut-like inhibitory motif for generating categorical
511 representations (Fig. 5C-E), and its ability to convert even linear response profiles to categorical
512 ones (Fig. 1E-inset: orange/brown vs. purple). The donut-like inhibitory motif may, therefore, be
513 a conserved neural circuit module for categorical selection and decision-making in general,

514 potentially playing a functional role in sculpting categorical responses that arise through learning
515 as well (Freedman and Assad, 2006; Xin et al., 2019).

516 Here, we described the donut-like motif in the context of selection across space, which is
517 underpinned by well-organized topographic maps of space. However, for many forms of
518 selection, such organized maps do not exist, with olfactory categorization being an extreme
519 example (Niessing and Friedrich, 2010), involving dense combinatorial coding of odors. The
520 findings from this study suggest that detailed investigations of the properties and connectivity of
521 inhibitory neurons in those cases may be a fruitful endeavor. In the subcortical circuit studied
522 here, which is conserved across all vertebrates (Knudsen, 2011; Wang et al., 2004), the long-
523 range suppression that is necessary for the donut-like motif is implemented by inhibitory neurons
524 with far reaching projections. Alternatively, it could also be implemented, for instance, in
525 cortical circuits, through long-range excitation contacting local inhibitory neurons. Additionally,
526 our results testing different implementations of the donut-like motif to match leading classes of
527 models of decision-making support the generality of this neural circuit motif for implementing
528 categorization across brain areas.

529 A seeming alternate mechanism for categorization that we did not examine in detail is the
530 biophysical specification of input-output functions of the neurons to be step-like. However, the
531 limited power of this mechanism for producing categorical responses in general circumstances,
532 together with the past demonstration that the steepness of the i/o functions of neurons can be
533 uncorrelated with whether selection is categorical (Mysore et al., 2011), made this alternative not
534 compelling. Another alternative are highly recurrent networks, which have been shown in
535 modeling to be capable of generating categorical outputs from multiplexed encoding through

dynamics (Chaisangmongkon et al., 2017; Mante et al., 2013). However, because it is difficult to extract specific, testable neural circuit mechanisms from such a network, and in light of the recently reported counterpoint to mixed-selectivity descriptions (Hirokawa et al., 2019), we focused, here, on structured circuit mechanisms. We considered explicitly the alternatives of feedback inhibition and recurrent amplification. Contrary to prior proposals (Mysore and Knudsen, 2012; Wang, 2008), feedback inhibition, by itself, was 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 enhanced 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 - the implementation of a flexible selection boundary (feedback inhibition; (Fadok et al., 2017; Jovanic et al., 2016; Mysore and Knudsen, 2012)), and evidence accumulation (recurrent amplification; (Wang, 2008, 2012)), selection circuits may need to include these motifs. Together with the donut-like inhibitory motif, they can then effectively implement attractor dynamics for flexibly categorical decision-making (Machens et al., 2005; Mysore and Knudsen, 2012; Niessing and Friedrich, 2010; Wang, 2008). An intriguing open question in this context, even in the avian midbrain selection network, is how the wiring of the exquisitely organized (multi-holed) donut-like connectivity is achieved in the brain.

558 **Acknowledgments**

559 This work was supported in part by funding from NIH R01 EY027718. NRM and SPM designed
560 the research, NRM performed experiments and analyzed the data, SPM and NRM wrote the
561 paper. The authors declare no competing interests.

562

563 ## Methods

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

570

571 **Neurophysiology.** Experiments were performed following protocols that have been described
572 previously (Mysore et al., 2010; Mysore and Knudsen, 2013). Briefly, epoxy-coated, high
573 impedance, tungsten microelectrodes (A-M Systems, 250μm, 5-10MΩ at 1 kHz) were used to
574 record single and multi-units extracellularly in the OTid. Multi-barrel glass electrodes (Kation
575 Scientific, Carbostar– 3LT, 0.4-1.2MΩ at 1kHz) filled with kynurenic acid (a competitive inhibitor
576 of ionotropic glutamate receptors; pH 8.5-9 at a concentration of 40mM) were used to record from
577 and inactivate neurons in the Imc and Ipc. Inactivation was performed using micro iontophoresis
578 by ejecting kynurenic acid with an eject current of -450 nA to -500 nA; data were collected starting

579 15 min after drug ejection commenced. A retain current of +15 nA was used to prevent leakage of
580 the drug from the tip of the electrodes when drug was not being iontophoresed. Recovery data
581 were measured 15 min after drug ejection was ceased.
582
583 **OT, Imc and Ipc targeting.** We navigated to the OT (based on well-established methods
584 (Knudsen, 1982)), and then navigated to the Imc using the OT's topographic space map as
585 reference. The Imc is an oblong structure that is 2.8mm rostrocaudally and 0.35mm dorsoventrally,
586 appearing as a 700-µm x 350-µm elliptical disk in coronal sections. It lies parallel to the
587 rostrocaudal axis of the OT, located approximately 16 mm ventral to the surface of the brain, and
588 approximately 500 µm medial to the medial-most part of the OT. We targeted the Imc following
589 published methods (Mysore and Knudsen, 2013) Imc targeting has been validated previously using
590 dye injections and lesions (Mahajan and Mysore, 2018). Dorsoventral penetrations through the
591 Imc were made at a medial-leading angle of 5˚ from the vertical to avoid a major blood vessel in
592 the path to the Imc. The Ipc lies roughly 500-700 um medial to the Imc and its targeting was
593 confirmed based on the neural response characteristics of the neurons (characteristic bursty
594 responses; (Asadollahi et al., 2010)).
595
596 **Data collection and spike sorting.** Multi-unit spike waveforms were recorded using Tucker Davis
597 Technologies hardware interfaced with MATLAB. The responses of neurons were measured by
598 counting spikes during a time window (typically 100-350 ms) following stimulus onset.
599 The automated 'wave-clus' spike-sorting toolbox was used for spike sorting (Chaure et al.,
600 2018). We included only those units for analysis for which fewer than 5% of the recorded spikes

601 were within 1.5 ms (inter-spike interval; ISI) of each other.

602

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

604

605 **Input output functions.** The input output functions (firing rate f, as a function of the saliency, l)
606 of the neurons in the model were modeled using sigmoid functions using previously published
607 methods(Mysore and Knudsen, 2012).

608 $$f(l) = c + s\left(\frac{l^m}{l^m + L_{50}^m}\right)$$

609

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

615

616 **Excitatory neurons**. The excitatory neurons in Figs. 1, S1, S2, and S4 were simulated using the
617 following parameters (which are consistent with the parameters obtained by fitting a sigmoid to
618 response functions of OTid neurons (Mysore and Knudsen, 2012):

619 $$c = 5.3, s = 22.2\,, L_{50} = 11.6, m = \ 2$$

620

621 **Inhibitory neurons.** The inhibitory neurons in Figs. 1, S1, S2, and S4 were simulated using the
622 following parameters (which are consistent with the parameters obtained by fitting a sigmoid to
623 response functions of Imc neurons (Mysore and Knudsen, 2012):

624 $$c = 5, s = 15\ , L_{50} = 8, m = 10$$

625 **Recurrent excitation neurons.** The excitatory neurons that provide recurrent amplification in Fig.
626 S2 were simulated using the following parameters (which are consistent with the parameters
627 obtained by fitting a sigmoid to response functions of Ipc neurons; (Asadollahi et al., 2011)):

628 $$c = 8.4, s = 36\ , L_{50} = 5.8, m = \ 3.3$$

629 The inhibition sent from the inhibitory neurons onto excitatory neurons is modeled using
630 input and output divisive factors as below using previously published methods (Mysore and
631 Knudsen, 2012).

632 $$f = \left(\frac{1}{s_{out}} + 1\right).\left(\frac{c}{s_{in}+1} + s\left(\frac{I^m}{I^m + L_{50}^m + s_{in}^m}\right)\right), \text{ where}$$

633 - $s_{in} = d_{in}.I, s_{out} = d_{out}.I$ are the input and output divisive factors.
634 - $d_{in}\ and\ d_{out}$ are parameters that control the strength of input and output division and I is
635 the output (firing rate) of the inhibitory neuron sending the inhibition.

636 The value of these parameters chosen were $d_{in} = 0\ and\ d_{out} = 0.25$ consistent with
637 previously published methods (see Fig. 5D in (Mysore and Knudsen, 2012)).

638

639 **Feedback inhibition.** In models in which the inhibitory neurons inhibit each other (e.g. Fig. 1C,
640 bottom left panel, blue connections; feedback model), the feedback inhibition was modeled as
641 below using previously published methods(Mysore and Knudsen, 2012).

642

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

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

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).

663 In models which do not have the donut-like connectivity (Fig. 1C, top left panel; Baseline model),
664 s is set to 1.

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

668

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

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

674

675 $$f(l) = (1 + e_{out}).\left(\frac{1}{s_{out} + 1}\right).(\frac{c}{s_{in} + 1} + s\left(\frac{l^m}{l^m + L_{50}^m + s_{in}^m}\right))$$

676

677 where, $e_{out} = e_a.A, \quad e_a = 0.01$, and A is the output of the neuron sending the amplification.
678 The value of the parameter is chosen as 0.01 to yield results consistent with the amplification effect
679 of Ipc on OT responses as reported in previously published work (Ipc inactivation results in a 31%
680 decrease in the OTid responses on average; (Asadollahi and Knudsen, 2016)).

681

682 **Models for Figs. 5 and 6.** To implement the models in Figs. 5 and 6 (normalization model and
683 donut-like motifs purely via a recurrent route), we used the input-output functions and the effect
684 of divisive inhibition described above. The primary feature of these models that is different from

models in Figs. 1, S1 and S2 is that the excitatory neurons send inputs to inhibitory neurons, which then inhibit those excitatory neurons. The output of an excitatory neuron at time t is calculated by applying divisive inhibition (from the inhibitory neurons at time t-1) to its activity at time t-1. This activity is then used to calculate the activity of the inhibitory neurons at time t as below.

$$I(t) = (1 + e_{out}).I(t-1)$$

where, $e_{out} = e_a.A, \quad e_a = 0.01$, and A is the output of the neuron sending the input at time t-1. This process is repeated iteratively until steady state activity is obtained. The initial activities (at t=0) of the inhibitory neurons are set to their baseline level, and of the excitatory neurons are calculated from the stimulus inputs without any divisive inhibition.

For the models with recurrent amplification (Fig. 6AB), the output of the neuron is scaled by a factor (k, k= 2.5).

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 (Mysore and Knudsen, 2011a). 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

707 response profiles (mean +/- s.e.m.) of that neuron from different model circuits. Similarly,
708 response profiles for 50 neurons were obtained for each model circuit. For all the model runs
709 reported in Figs. 1, S1, S2 and 5, we used a fano factor value of 6. We also varied the fano factor
710 value to test the effect of noise on model performance (Fig. S1D).

711 This stimulus presentation protocol results in 2 categories (Category 1: S1 > S2 and
712 Category 2: S1 < S2). We measured the boundary discriminability (bd') of the responses of neuron
713 1 by calculating the d-prime between the responses of neuron to stimuli pair straddling either side
714 of the selection boundary and at a distance of 3 units from the boundary as:

715

716 $$bd' = \frac{\mu_1 - \mu_2}{\sqrt{0.5(s_1^2 + s_2^2)}}$$

717

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

721 To compute the categorization index, we compared two quantities: (modified from
722 previously published work; (Freedman and Assad, 2006; Mysore and Knudsen, 2011a)) (a) the
723 mean within-category d-prime (WCD') between the responses of the neuron to pairs of stimulus-
724 pairs (S1 and S2) that are in the same category, and (b) the mean between-category d-prime (BCD')
725 between the responses of the neuron to pairs of stimulus-pairs that are in different categories. The
726 pairs are chosen while ensuring that (i) the number of pairs used to calculate both these metrics
727 are the same, and (ii) the distribution of the distances between the chosen pairs for calculating both
728 the metrics are the same. The categorization index is calculated from these two metrics as:

729 $$\text{Categorization index (CatI)} = \frac{\text{mean}(\text{BCD}') - \text{mean}(\text{WCD}')}{\text{mean}(\text{BCD}') + \text{mean}(\text{WCD}')}$$

730

731 CatI= 1 indicates idealized, step-like responses; = 0 indicates linear, non-categorical responses;

732 <0 indicates better discriminability within category than between categories.

733

734 **Data collection protocol.**

735 Visual stimuli used here have been described previously (Mysore et al., 2010; Mysore and

736 Knudsen, 2011a). Briefly, looming visual dots are flashed at different locations on a tangent TV

737 monitor in front of the owl. Looming stimuli were dots that expanded linearly in size over time,

738 starting from a size of 0.6˚ in radius. The speed of the loom was decided based on the stimulus

739 protocol and varied between 9.6°/s and 21.6°/s. Visual stimuli were presented for a duration of

740 250ms with an inter stimulus interval of 1.25s to 4s.

741

742 **Receptive fields (RFs).** For measuring spatial RFs of Imc, OTid and Ipc neurons, a single stimulus

743 of a fixed contrast was presented at the sampled locations. The order of locations at which the

744 stimulus was presented was randomized to minimize adaptation. A location is considered to be

745 inside the RF if it evokes responses significantly different from that of baseline response. All other

746 locations are considered to be outside the RF.

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

748

749 **I. Paired OTid and Imc data collection**

750 1. *"Other" inhibition*

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

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

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

760 b) Tuning curve centered at the OTid RF peak + competing stimulus centered at the Imc RF
761 peak (TCC): The same curve as in a), but along with a (distant) competitor positioned at
762 the peak of the Imc RF. The strength of the competing stimulus was chosen to be between
763 19.2°/s. This second competing stimulus drives the Imc neuron, which sends strong
764 competitive inhibition to the OTid neuron. Since the competing stimulus was more salient
765 than the stimulus driving the OTid neuron, the responses of the OTid neuron were strongly
766 suppressed consistent with previous published results (Mysore et al., 2010).

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

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

773 min after the drug ejection was started. Recovery curves were measured 15 min after the drug
774 ejection was stopped.
775 We analyzed the data from these three curves as below.
776 First, a distant competing stimulus is known to typically suppress OTid responses (Mysore
777 et al., 2010; Rizzolatti et al., 1974), something that we confirmed by comparing the responses of
778 OTid neurons to the TC (curve a)) and TCC (curve b)); Fig. S2E. Notably, it is also known that
779 some OTid neurons do not show such suppression, and any such neurons were excluded from our
780 analyses (Mysore et al., 2011).
781 Next, to measure the amount of Imc inactivation, we compared the responses of the Imc
782 neurons to the TC measured at the Imc RF peak (curve c) in the baseline condition versus the
783 inactivation condition. Kynurenic acid was able to effectively shut down Imc responses: (Fig 2F;
784 red, median strength of Imc inactivation = 95%, 95% CI = [87%, 103%], p = 3.8e-6, sign test,
785 n=19).
786 Finally, to measure the strength of "other" inhibition, we compared the TCC responses (curve
787 b) in the baseline condition and the inactivation condition. In the baseline condition, the Imc
788 neuron was intact, driven by the competing stimulus and sent strong inhibition to the OTid neuron.
789 In the inactivation condition, the Imc neuron was shut off, as the result of which the OTid neuron
790 was released from inhibition and exhibited an increase in responses. The difference between these
791 two curves quantified the strength of "other" inhibition from the Imc onto OTid.
792
793 2. *"Self" inhibition*:
794 To measure the strength of "self" inhibition sent from an Imc neuron to a matched location in the

795 OTid space map, we simultaneously recorded from the Imc neuron (with a multi-barrel glass
796 electrode filled with the kynurenic acid), and a spatially aligned "self" OTid neuron (with a
797 tungsten electrode) as described below; we ensured that the half-max of the RF of the OTid neuron
798 overlapped with the half-max of the Imc RF.

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

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

804 To measure the amount of Imc inactivation, we compared the responses of the Imc neuron in
805 the baseline and the inactivation condition as before. kynurenic acid was able to effectively shut
806 down Imc responses: (median strength of Imc inactivation = 92%, 95% CI = [86% 98%], p = 7.5e-
807 9, sign test).

808

809 3. *"Gap" inhibition*:

810 To measure the strength of "self" inhibition sent from an Imc neuron with a multilobed RF to the
811 locations in the OTid space map between the RF lobes (gaps), we simultaneously recorded from
812 the Imc neuron (with a multi-barrel glass electrode filled with the kynurenic acid), and "gap" OTid
813 neuron (with a tungsten electrode), the RF (half-max) of which was located in the gap between the
814 half-max of the lobes of the Imc RF (Fig. 3A-C). We recorded the same set of curves as in the
815 "other" inhibition case (see above) from the gap OTid neuron, and performed similar analyses on
816 the data to quantify the strength of "gap" inhibition.

817 4. *"Different lobe" inhibition*:

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

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

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

838

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 (Mysore and Knudsen, 2011a). 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.

Data analyses and statistical tests. 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).

Code and data availability. Software code and the data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Albantakis, L., and Deco, G. (2009). The encoding of alternatives in multiple-choice decision making. Proc Natl Acad Sci U S A *106*, 10308-10313.

Asadollahi, A., and Knudsen, E.I. (2016). Spatially precise visual gain control mediated by a cholinergic circuit in the midbrain attention network. Nature communications *7*, 13472.

Asadollahi, A., Mysore, S.P., and Knudsen, E.I. (2010). Stimulus-driven competition in a cholinergic midbrain nucleus. Nat Neurosci *13*, 889-895.

Asadollahi, A., Mysore, S.P., and Knudsen, E.I. (2011). Rules of competitive stimulus selection in a cholinergic isthmic nucleus of the owl midbrain. J Neurosci *31*, 6088-6097.

Bathellier, B., Ushakova, L., and Rumpel, S. (2012). Discrete neocortical dynamics predict behavioral categorization of sounds. Neuron *76*, 435-449.

Bollimunta, A., and Ditterich, J. (2012). Local computation of decision-relevant net sensory evidence in parietal cortex. Cereb Cortex *22*, 903-917.

Busse, L., Wade, A.R., and Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. Neuron *64*, 931-942.

Carandini, M., Heeger, D.J., and Movshon, J.A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci *17*, 8621-8644.

Chaisangmongkon, W., Swaminathan, S.K., Freedman, D.J., and Wang, X.J. (2017). Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential, Category-Based Decisions. Neuron *93*, 1504-1517 e1504.

Chaure, F.J., Rey, H.G., and Quian Quiroga, R. (2018). A novel and fully automatic spike-sorting implementation with variable number of features. J Neurophysiol *120*, 1859-1871.

883 Chen, Q., Pei, Z., Koren, D., and Wei, W. (2016). Stimulus-dependent recruitment of lateral
884 inhibition underlies retinal direction selectivity. eLife *5*.

885 Churchland, A.K., and Ditterich, J. (2012). New advances in understanding decisions among
886 multiple alternatives. Current Opinion in Neurobiology *22*, 920-926.

887 Engel, T.A., Chaisangmongkon, W., Freedman, D.J., and Wang, X.J. (2015). Choice-correlated
888 activity fluctuations underlie learning of neuronal category representation. Nature
889 communications *6*, 6454.

890 Fadok, J.P., Krabbe, S., Markovic, M., Courtin, J., Xu, C., Massi, L., Botta, P., Bylund, K.,
891 Muller, C., Kovacevic, A.*, et al.* (2017). A competitive inhibitory circuit for selection of active
892 and passive fear responses. Nature *542*, 96-100.

893 Fecteau, J.H., and Munoz, D.P. (2006). Salience, relevance, and firing: a priority map for target
894 selection. Trends Cogn Sci *10*, 382-390.

895 Freedman, D.J., and Assad, J.A. (2006). Experience-dependent representation of visual
896 categories in parietal cortex. Nature *443*, 85-88.

897 Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical representation
898 of visual stimuli in the primate prefrontal cortex. Science *291*, 312-316.

899 Goddard, C.A., Mysore, S.P., Bryant, A.S., Huguenard, J.R., and Knudsen, E.I. (2014). Spatially
900 Reciprocal Inhibition of Inhibition within a Stimulus Selection Network in the Avian Midbrain.
901 PLoS One *9*, e85865.

902 Hartline, H.K., Wagner, H.G., and Ratliff, F. (1956). Inhibition in the eye of Limulus. J Gen
903 Physiol *39*, 651-673.

904 Hirokawa, J., Vaughan, A., Masset, P., Ott, T., and Kepecs, A. (2019). Frontal cortex neuron
905 types categorically encode single decision variables. Nature *576*, 446-451.

906 Jiang, X., Chevillet, M.A., Rauschecker, J.P., and Riesenhuber, M. (2018). Training Humans to
907 Categorize Monkey Calls: Auditory Feature- and Category-Selective Neural Tuning Changes.
908 Neuron *98*, 405-416 e404.

909 Jovanic, T., Schneider-Mizell, C.M., Shao, M., Masson, J.B., Denisov, G., Fetter, R.D., Mensh,
910 B.D., Truman, J.W., Cardona, A., and Zlatic, M. (2016). Competitive Disinhibition Mediates
911 Behavioral Choice and Sequences in Drosophila. Cell *167*, 858-870 e819.

912 Knudsen, E.I. (1982). Auditory and visual maps of space in the optic tectum of the owl. J
913 Neurosci *2*, 1177-1194.

914 Knudsen, E.I. (2011). Control from below: the role of a midbrain network in spatial attention.
915 Eur J Neurosci *33*, 1961-1972.

916 Kouh, M., and Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations.
917 Neural Comput *20*, 1427-1451.

918 Louie, K., Khaw, M.W., and Glimcher, P.W. (2013). Normalization is a general neural
919 mechanism for context-dependent decision making. Proc Natl Acad Sci U S A *110*, 6139-6144.

920 Lovejoy, L.P., and Krauzlis, R.J. (2010). Inactivation of primate superior colliculus impairs
921 covert selection of signals for perceptual judgments. Nat Neurosci *13*, 261-266.

922 Machens, C.K., Romo, R., and Brody, C.D. (2005). Flexible control of mutual inhibition: a
923 neural model of two-interval discrimination. Science *307*, 1121-1124.

924 Mahajan, N.R., and Mysore, S.P. (2018). Combinatorial Neural Inhibition for Stimulus Selection
925 across Space. Cell reports *25*, 1158-1170 e1159.

926 Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent
927 computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78-84.

928 Marin, G., Mpodozis, J., Sentis, E., Ossandon, T., and Letelier, J.C. (2005). Oscillatory bursts in
929 the optic tectum of birds represent re-entrant signals from the nucleus isthmi pars parvocellularis.
930 J Neurosci *25*, 7081-7089.

931 Marin, G., Salas, C., Sentis, E., Rojas, X., Letelier, J.C., and Mpodozis, J. (2007). A cholinergic
932 gating mechanism controlled by competitive interactions in the optic tectum of the pigeon. J
933 Neurosci *27*, 8112-8121.

934 McPeek, R.M., and Keller, E.L. (2004). Deficits in saccade target selection after inactivation of
935 superior colliculus. Nat Neurosci *7*, 757-763.

936 Mysore, S.P., Asadollahi, A., and Knudsen, E.I. (2010). Global inhibition and stimulus
937 competition in the owl optic tectum. J Neurosci *30*, 1727-1738.

938 Mysore, S.P., Asadollahi, A., and Knudsen, E.I. (2011). Signaling of the strongest stimulus in the
939 owl optic tectum. J Neurosci *31*, 5186-5196.

940 Mysore, S.P., and Knudsen, E.I. (2011a). Flexible categorization of relative stimulus strength by
941 the optic tectum. J Neurosci *31*, 7745-7752.

942 Mysore, S.P., and Knudsen, E.I. (2011b). The role of a midbrain network in competitive stimulus
943 selection. Curr Opin Neurobiol *21*, 653-660.

944 Mysore, S.P., and Knudsen, E.I. (2012). Reciprocal inhibition of inhibition: a circuit motif for
945 flexible categorization in stimulus selection. Neuron *73*, 193-205.

946 Mysore, S.P., and Knudsen, E.I. (2013). A shared inhibitory circuit for both exogenous and
947 endogenous control of stimulus selection. Nat Neurosci *16*, 473-478.

948 Mysore, S.P., and Knudsen, E.I. (2014). Descending control of neural bias and selectivity in a
949 spatial attention network: rules and mechanisms. Neuron *84*, 214-226.

950 Mysore, S.P., and Kothari, N.B. Mechanisms of competitive selection: A canonical neural circuit
951 framework In review

Niessing, J., and Friedrich, R.W. (2010). Olfactory pattern classification by discrete neuronal network states. Nature *465*, 47-52.

Olsen, S.R., Bhandawat, V., and Wilson, R.I. (2010). Divisive normalization in olfactory population codes. Neuron *66*, 287-299.

Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput *20*, 873-922.

Rizzolatti, G., Camarda, R., Grupp, L.A., and Pisa, M. (1974). Inhibitory effect of remote visual stimuli on visual responses of cat superior colliculus: spatial and temporal factors. J Neurophysiol *37*, 1262-1275.

Rodriguez, A., and Laio, A. (2014). Machine learning. Clustering by fast search and find of density peaks. Science *344*, 1492-1496.

Roe, R.M., Busemeyer, J.R., and Townsend, J.T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. Psychological Review *108*, 370-392.

Sereno, M.I., and Ulinski, P.S. (1987). Caudal topographic nucleus isthmi and the rostral nontopographic nucleus isthmi in the turtle, Pseudemys scripta. J Comp Neurol *261*, 319-346.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a dataset via the gap statistic. J R Statist Soc B *63*, 411-423.

Usher, M., and McClelland, J.L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev *108*, 550-592.

Wang, X.J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron *36*, 955-968.

Wang, X.J. (2008). Decision making in recurrent neuronal circuits. Neuron *60*, 215-234.

Wang, X.J. (2012). Neural dynamics and circuit mechanisms of decision-making. Curr Opin Neurobiol *22*, 1039-1046.

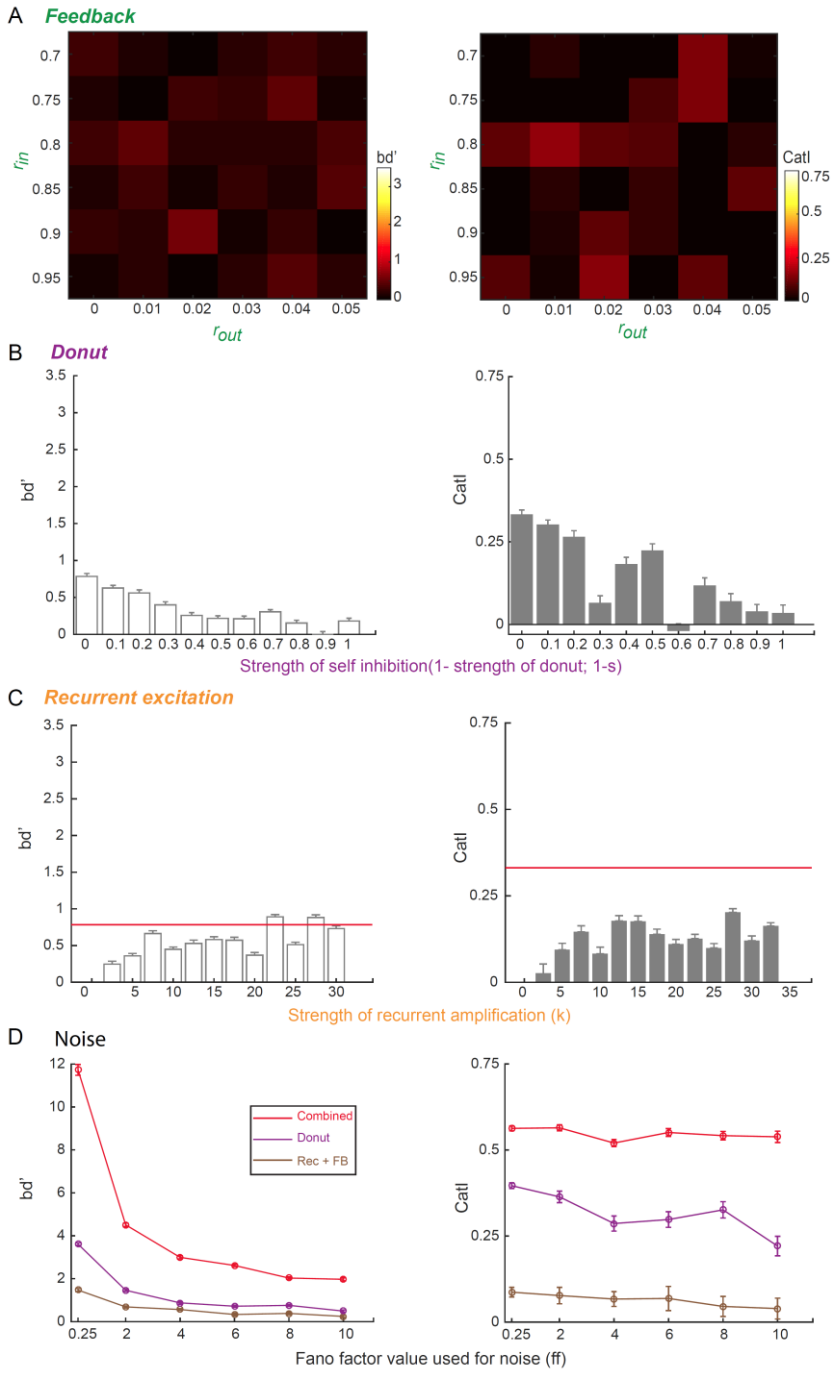
Wang, Y., Luksch, H., Brecha, N.C., and Karten, H.J. (2006). Columnar projections from the cholinergic nucleus isthmi to the optic tectum in chicks (Gallus gallus): a possible substrate for synchronizing tectal channels. J Comp Neurol *494*, 7-35.

Wang, Y., Major, D.E., and Karten, H.J. (2004). Morphology and connections of nucleus isthmi pars magnocellularis in chicks (Gallus gallus). J Comp Neurol *469*, 275-297.

Wong, K.F., and Wang, X.J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J Neurosci *26*, 1314-1328.

Xin, Y., Zhong, L., Zhang, Y., Zhou, T., Pan, J., and Xu, N.L. (2019). Sensory-to-Category Transformation via Dynamic Reorganization of Ensemble Structures in Mouse Auditory Cortex. Neuron *103*, 909-921 e906.

1001 **Supplemental Figures**



1002

1003 ***Fig. S1. Donut-like inhibition surpasses other circuit motifs in its ability to generate***

1004 ***categorical representations over a large range of values of key model parameters and amounts***

1005 ***of response noise (modeling). (A)*** *Effect of varying strength of feedback inhibition in model in*
1006 *Fig. 1C-bottom-left, on bd' (left) and CatI (right).* r_{in}*: input divisive factor;* r_{out}*: output divisive*
1007 *factor (Methods). Range of variation based on previously published work (Mysore and Knudsen,*
1008 *2012). Maximum bd' over the range of parameters is = 0.5478 (lower than from the donut-like*
1009 *motif, bd' = 0.785); maximum CatI = 0.157 (lower than from the donut-like motif, CatI = 0.331).*
1010 ***(B)*** *Effect of varying strength of "self" inhibition in model in Fig. 1C-top-right on bd' (left) and*
1011 *CatI (right). Left: Corr = -0.89, p = 1.64e-4 (correlation test). Right: Corr = -0.81, p = 3e-3*
1012 *(correlation test). Maximal effects on bd' and CatI are when "self" inhibition=0, i.e., when the*
1013 *circuit has donut-like inhibition.* ***(C)*** *Effect of varying strength of recurrent amplification in*
1014 *model in Fig. 1D-top-left, on bd' (left) and CatI (right). Red lines: values from circuit with*
1015 *donut-like motif only. Left: bd' increases with amplification (corr=0.68; p = 1e-3; corr test),*
1016 *saturating around k=10; the increase is as expected because of the progressive scaling up of*
1017 *responses due to amplification. Note that physiological range is under 3 (Asadollahi and*
1018 *Knudsen, 2016). Right: However, CatI does not change systematically (corr= -0.14; p =0.54;*
1019 *corr test), indicating that varying the strength of amplification either does not affect the*
1020 *estimates in Fig. 1F for this motif.* ***(D)*** *Comparison across three models of bd' (left) and CatI*
1021 *(right) from simulated response profiles as a function of fano-factor. Models are circuit with*
1022 *donut-like inhibitory motif only (Fig. 1C, top-right), or circuit with recurrent amplification and*
1023 *feedback inhibition together (Fig. 1D, bottom-left), or circuit with all three motifs combined*
1024 *(Fig. 1D, bottom-right).*

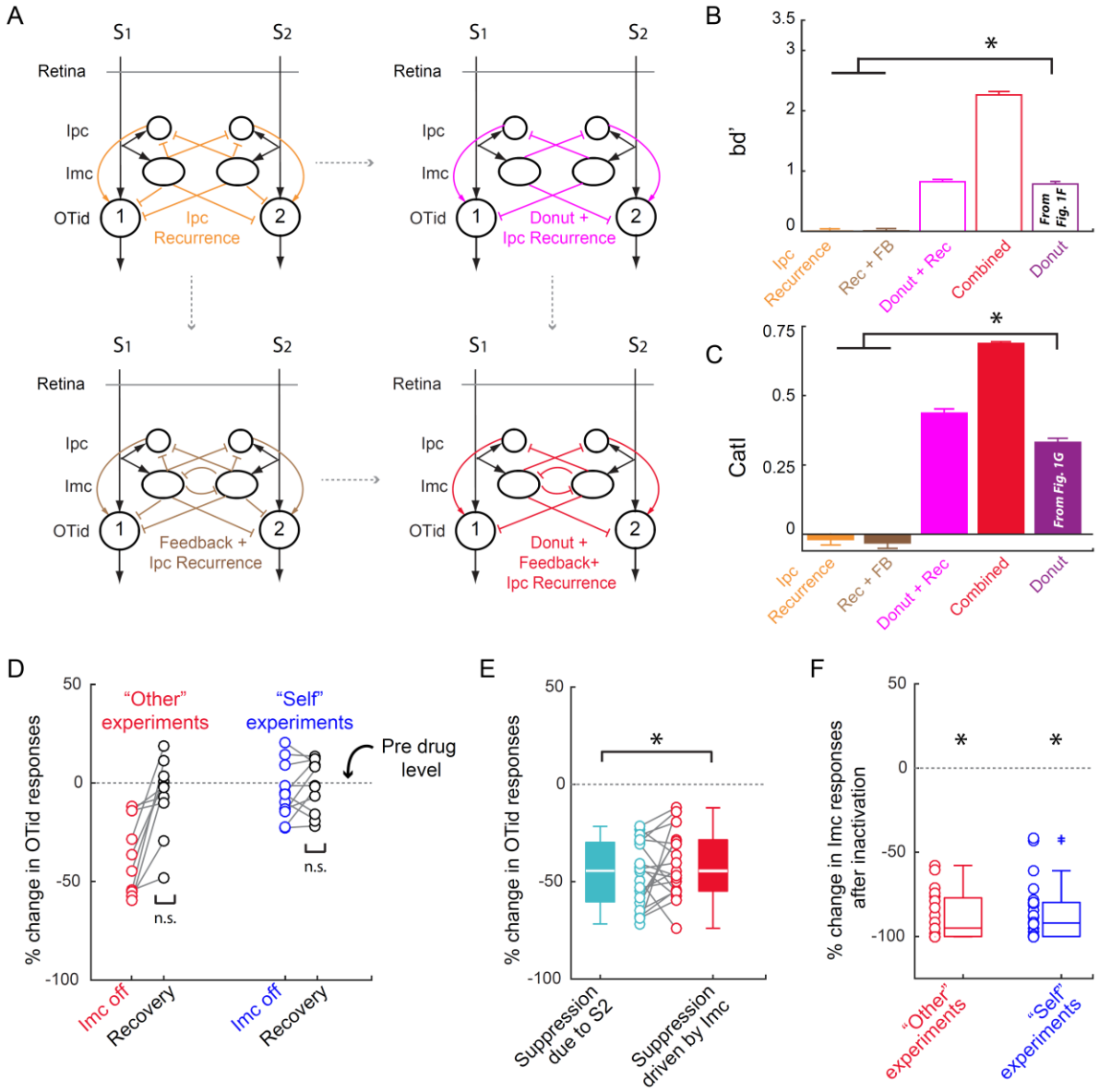


Fig. S2. Donut-like inhibition surpasses other circuit motifs even in the case of biologically grounded model (model based on barn owl Imc-Ipc-OT connectivity and function), and supporting data for Fig. 2 (experiments). (A) *Baseline + recurrent excitation model, similar to the one in Fig. 1D, top-left, 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"*

1032 *inhibition to zero; this is done both for the direct inhibitory pathway from Imc → OTid, and the*
1033 *indirect pathway: Imc → Ipc → OTid. Bottom-right: Adding both feedback inhibition and donut-*
1034 *like inhibition.* ***(B)*** *bd' for these four models in A, computed on the response profiles of model*
1035 *neuron 1 obtained by simulating with the same two-stimulus morphing protocol used in Fig. 1E.*
1036 *Shown ghosted in for comparison are the results from the circuit model in Fig. 1C, top-right,*
1037 *that has just the donut-like motif without any recurrent amplification or feedback inhibition. '*':*
1038 *p = 9.5e-10, ANOVA followed by HBMC correction for multiple comparisons.* ***(C)*** *Same as B,*
1039 *but CatI. Results show that the supremacy of the donut-like motif over recurrent amplification*
1040 *alone, or the combination of that with feedback inhibition, in generating categorical response*
1041 *profiles still holds in the (biologically grounded) model of the barn owl midbrain circuit. '*': p =*
1042 *9.6e-10, ANOVA followed by HBMC correction for multiple comparisons.* ***(D)*** *Recovery of OTid*
1043 *responses from kynurenic acid iontophoresis for experiments in Fig. 2. Data on left: "Other"*
1044 *experiments. OTid responses revealed significant change during iontophoresis (red dots; subset*
1045 *of the data reproduced from Fig. 2L), but returned to pre-drug baseline (horizontal line) in*
1046 *recovery (black dots; p=0.32, t-test against 0). Recovery data obtained 15 min after*
1047 *iontophoretic eject current was switched to retain current; Methods). Data show recovery,*
1048 *demonstrating that the effects reported Fig. 2 are due specifically to drug iontophoresis/Imc*
1049 *inactivation. Data on right: "Self" experiments. OTid responses showed no significant change*
1050 *during iontophoresis (blue dots; reproduced from Fig. 2L), and stayed around zero in recovery*
1051 *(black dots; p=0.58, t-test against 0.* ***(E)*** *"Other" experiment. Comparison of suppression*
1052 *provided by Imc with that due to S2 (i.e., the maximum amount of suppression experienced by the*
1053 *OTid neuron in this stimulus protocol). This is done by comparing (i) % suppression of OTid*

1054 *responses by stimulus S2 when Imc is intact, computed by comparing OTid responses to S1 alone*
1055 *versus the paired presentation of S1 and S2 (teal;(Mysore et al., 2010)), with (ii) % suppression*
1056 *of OTid responses to paired presentation of S1 and S2 produced by Imc inactivation, computed*
1057 *by comparing OTid responses in the Imc-intact to Imc-off conditions (red; data reproduced from*
1058 *Fig. 2L). Consistent with (Mysore and Knudsen, 2013), nearly all the suppression due to*
1059 *competitor S2 (teal) is supplied by Imc (red) (p= 0.22, t-test, teal vs. red), verifying that Imc*
1060 *supplies powerful "other" inhibition.* ***(F)*** *Quantifying the effectiveness of Imc inactivation by*
1061 *kynurenic acid iontophoresis in the "other" (red) and "self"-inhibition (blue) experiments. In*
1062 *both cases, the inactivation was highly effective. Red: median % change in response; median =*
1063 *95%, 95% CI = [87%, 103%], p = 3.8 e-6, sign test; Blue: median % change in responses*
1064 *median =0.92, 95% CI = [86%, 98%], p = 7.45e-9, sign test.*

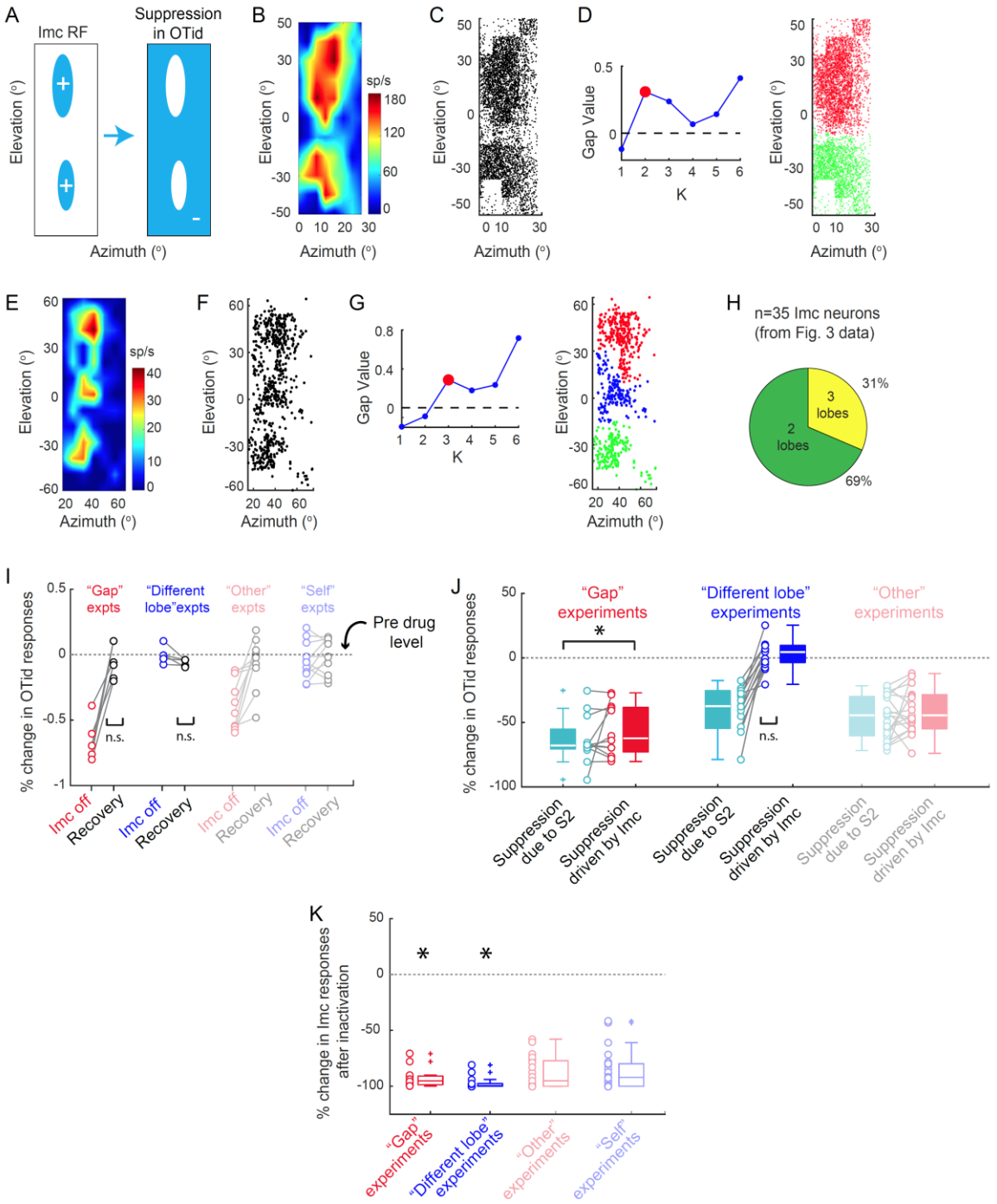


Fig. S3. Detecting the number of Imc RF lobes, and other supporting experimental data for Figure 3. (A) *Schematic representation of a multi-holed donut. Left: Two-lobed RF of a putative Imc neuron (left; blue ovals are the RF lobes). Right: Pattern of inhibition created by this neuron*

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 (Mahajan and Mysore, 2018).* ***(D)*** *Left: Density peaks clustering method (Rodriguez and Laio, 2014) 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 (Tibshirani et al., 2001) 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 (Mysore and Knudsen, 2013), nearly all the suppression due to*

competitor S2 (teal) is supplied by Imc (red) (p= 0.47, teal vs. red, ranksum test). "Different lobe" expt: 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 (Mahajan and Mysore, 2018), 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.*

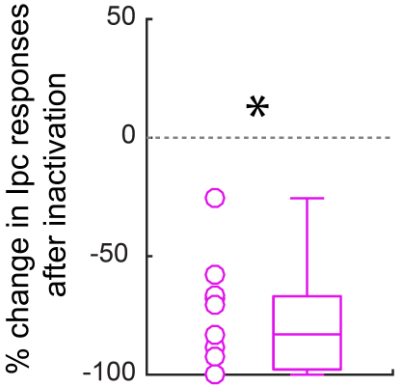


Fig. S4. Ipc is effectively inactivated (experiments re. Figure 4). *Quantifying the effectiveness of Ipc inactivation by kynurenic acid iontophoresis in the "necessity of donut-like inhibition" experiment. The inactivation was highly effective: (mean % change in responses = -77.42%, sd =22.82% p =5.3e-7, t-test against 0).*