bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The role of frontal cortex in multisensory decisions

Philip Coen[1*], Timothy P.H. Sit[2], Miles J Wells[1], Matteo Carandini[3], Kenneth D Harris[1]
[1]UCL Queen Square Institute of Neurology, University College London, London, United Kingdom
[2]Sainsbury-Wellcome Centre, University College London, London, UK
[3]UCL Institute of Ophthalmology, University College London, London, United Kingdom
*Correspondence: p.coen@ucl.ac.uk

Abstract

To interpret the world and make accurate perceptual decisions, the brain must combine information across sensory modalities. For instance, it must combine vision and hearing to localize objects based on their image and sound. Probability theory suggests that evidence from multiple independent cues should be combined additively, but it is unclear whether mice and other mammals do this, and the cortical substrates of multisensory integration are uncertain. Here we show that to localize a stimulus, mice combine auditory and visual spatial cues additively, a computation supported by unisensory processing in auditory and visual cortex and additive multisensory integration in frontal cortex. We developed an audiovisual localization task where mice turn a wheel to indicate the joint position of an image and a sound. Scanning optogenetic inactivation of dorsal cortex showed that auditory and visual areas contribute unisensory information, whereas frontal cortex (secondary motor area, MOs) contributes multisensory information to the decision of the mouse. Neuropixels recordings of >10,000 neurons in frontal cortex indicated that neural activity in MOs reflects an additive combination of visual and auditory signals. An accumulator model applied to the sensory representations of MOs neurons reproduced behaviourally observed choices and reaction times. This suggests that MOs integrates information from multiple sensory cortices, providing a signal that is then transformed into a binary decision by a downstream accumulator.

Introduction

The ability to combine visual and auditory information to better localize objects in our environment is critical to many species – whether prey, predator, or pedestrian crossing the street – and the optimal way to combine it is to add functions summarizing the evidence from each modality. By the laws of probability theory, optimally integrating information from multiple sources with independent noise requires adding functions summarizing the evidence from each source[1]. For instance, to determine whether a stimulus is on the right or left, given visual evidence V and auditory evidence A, the log odds of a stimulus being on the right vs. the left is a sum of a visual-only and an auditory-only function (Appendix 1):

$$\log\left(\frac{p(R \mid A,V)}{p(L \mid A,V)}\right) = \log\left(\frac{p(V \mid R)}{p(V \mid L)}\right) + \log\left(\frac{p(A \mid R)}{p(A \mid L)}\right) + \log\left(\frac{p(R)}{p(L)}\right) = f(V) + g(A) + b \qquad (1)$$

In this equation, the functions $f(V)$ and $g(A)$ quantifying the evidence from each modality may be non-linear, but they are added together linearly. Additive integration is consistent with many observations from humans, primates, and rats[2–9], but other studies suggest that humans[10–12], primates[13], and mice[14] may break it through a bias toward a dominant sensory modality.

The roles of different cortical regions in multisensory behaviour, and the neural code they use to represent multisensory signals, is debated. To conclude a brain region is involved in multisensory integration, it must not only be shown to contain neurons encoding information from both sensory modalities, but also to have a causal role in behavioural responses to both modalities, alone or in combination. In rodents and other mammals, including humans, several brain regions appear to encode multiple modalities, including superior colliculus[15–19], thalamus[20–23], parietal cortex[3,5,6,14,24–32], frontal cortex[33–35], and even primary sensory cortices[36–45]. However, the causal role of these regions in multisensory decisions remains unclear. Perturbation studies have focused primarily on parietal cortex, and most[24,46,47], though not all[14], of these studies did not find a critical role for parietal cortex in multisensory behaviour. Thus, the question of which cortical areas, if any, are required for multisensory decisions, and the nature of the multisensory neural code in these regions, remains open[48].

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

To address these questions, we developed a two-alternative forced choice audiovisual spatial localization task for mice (Fig. 1a). We extended a previous visual detection task, where mice turn a steering wheel with their forepaws to indicate the position of a visual grating[49], by adding a semi-circular array of speakers. On each trial, mice were presented with a flashing visual grating on the left or right screen with variable contrast (including zero contrast, i.e. a grey screen), and a synchronous amplitude-modulated noise played from the left, centre, or right speaker. On coherent multisensory trials (where auditory and visual stimuli were on the same side), and on unisensory trials (where either visual contrast was zero or the auditory stimulus was central), mice earned a water reward for indicating the side where the stimuli were presented. On conflict multisensory trials (auditory and visual stimuli on different sides), or neutral trials (central auditory and zero contrast visual), mice were rewarded randomly (Extended Data Fig. 1a). Mice learned to perform this task proficiently, reaching 96% $\pm$ 3% correct (s.d., n = 17 mice) for the easiest stimuli (coherent trials with the highest visual contrast).

Mice performed this task more rapidly when the cues were coherent, suggesting that the task succeeded in engaging a multisensory decision process rather than separate unisensory processes (Fig. 1b). Reaction times varied across mice, with a median of 190 $\pm$ 120 ms (median $\pm$ m.a.d., n = 156k trials in 17 mice), and varied smoothly with stimulus contrast (Extended Data Fig. 1h). In unisensory auditory trials, reaction times were 22 $\pm$ 20 ms faster than in unisensory visual trials (s.d., n = 17 mice, $p < 0.001$, paired t-test), suggesting that the circuits responsible for audiovisual decisions receive auditory spatial signals earlier than visual signals[50,51] (Fig. 1b, Extended Data Fig. 1j,l). In multisensory trials, reaction times were 25 $\pm$ 18 ms faster for coherent than conflict trials ($p < 0.001$, paired t-test), consistent with coactivation of a multisensory decision process rather than with two unisensory processes racing each other to get to a decision[4,52].

Spatial localization task reveals additive audiovisual integration

Mice used both modalities to perform the task, even when the two were in conflict (Fig. 1c-e). The fraction of rightward choices depended gradually on visual stimulus contrast (*grey* psychometric curve), and was markedly increased when sounds were presented on the right (*red*) and decreased when they were presented on the left (*blue*). Mice performed more accurately on coherent trials than unisensory trials of either modality (Fig. 1e, Extended Data Fig. 1i,k), indicating that they were attending to both modalities[7,8,50,51,53,54].

To test whether mice make multisensory decisions according to the additive law, we fit an additive model to their behavioural responses. Equation 1 can be rewritten as:

$$p(R) = \sigma(f(V) + g(A) + b) \qquad (2)$$

where $p(R)$ is the probability of making a rightward choice, and $\sigma(x) = 1/(1 + exp(-x))$ is the logistic function. For the functions of the visual and auditory inputs we chose simple expressions:

$$f(V) = v_R {V_R}^{\gamma} - v_L {V_L}^{\gamma} \qquad (3)$$

$$g(A) = a_R A_R - a_L A_L \qquad (4)$$

where V_R and V_L are right and left visual contrasts (at least one of which was always zero), and A_R and A_L are indicator variables for right and left auditory stimuli (with value 1 or 0 depending on the position of the auditory stimulus). The visual contrast was raised to a power γ to account for contrast saturation in the visual system[55]. The model thus has 6 free parameters: the bias b, the visual exponent γ, two visual sensitivities v_R and v_L, and two auditory sensitivities a_R and a_L. This parameterisation is slightly redundant for symmetrical observers such as the 17 mice in Fig. 1, where it would suffice to have a single visual sensitivity and a single auditory sensitivity, but it will come in useful to model the effects of brain inactivations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

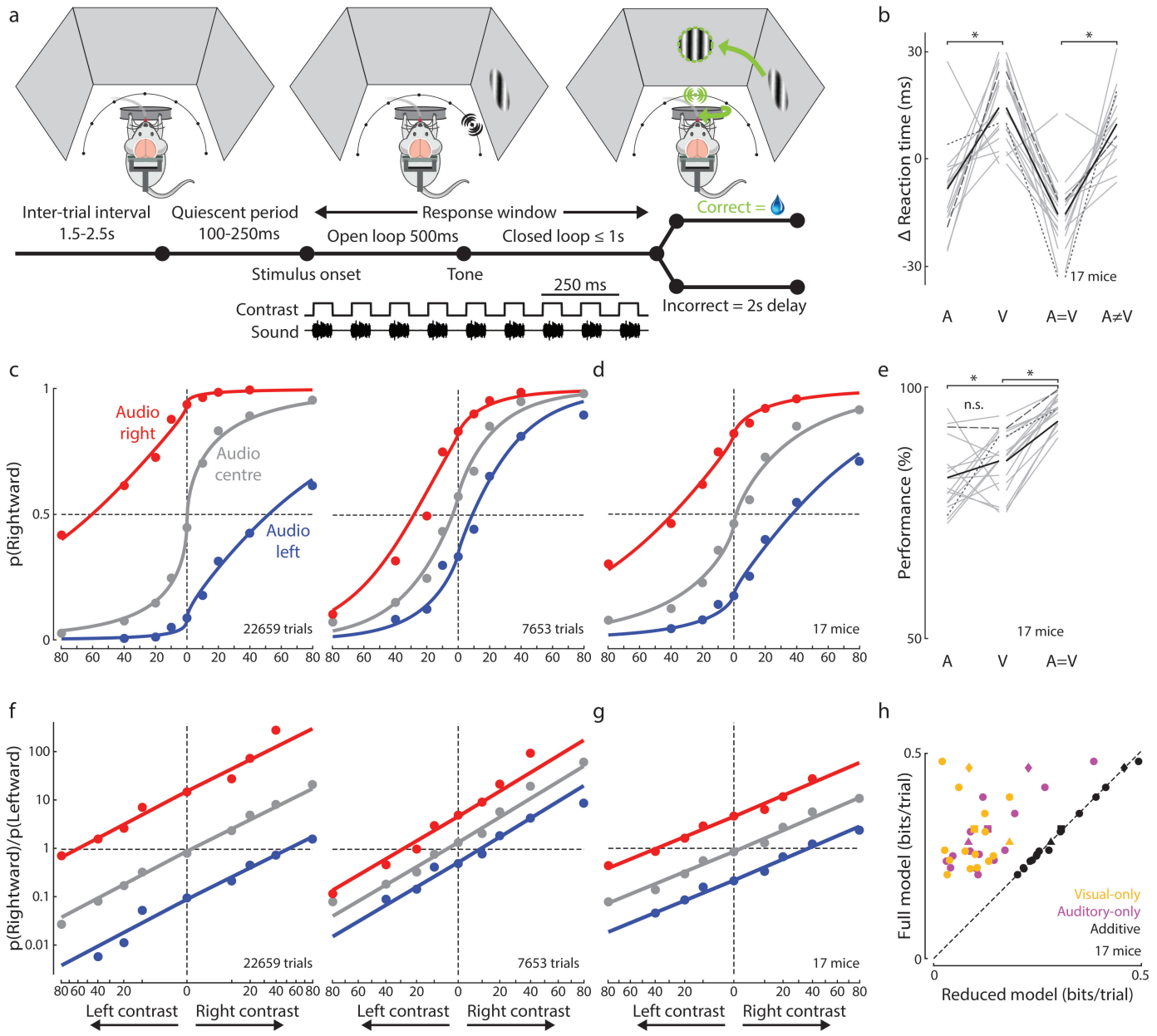


Fig. 1 Spatial localization task reveals additive audiovisual integration

(a) *Behavioural task. Visual and auditory stimuli are presented using 3 screens and an array of 7 speakers. In the example shown, auditory and visual stimuli are both presented on the right, and the subject is rewarded for turning the wheel anti-clockwise to bring the stimuli to the centre (a "rightward choice"). Bottom: Task timeline. After an inter-trial interval of 1.5-2.5 s, there is a 100-250 ms quiescent period where mice must hold the wheel still before stimuli appear. They then have 1.5 s to indicate their choice. After training, over 90% of choices were initiated within the "open loop" period (500 ms, Extended Data Fig. 1d-f), during which wheel movements did not affect stimulus location.* ***(b)*** *Median reaction times for each stimulus type relative to the mean across stimulus types for each mouse. Only trials with 40% visual contrast were included. Grey and black lines indicate individual mice and the mean across 17 mice. Long and short dashes indicate example mice from the left and right of (c). $*$: $p < 0.001$ (paired t-test).* ***(c)*** *The fraction of rightward choices at each visual contrast and auditory stimulus location for two example mice. Curves: fit by the additive behavioural model.* ***(d)*** *As in (c), but averaged across 17 mice (~156K trials). Curves: combined fit across all mice.* ***(e)*** *Mouse performance (% rewarded trials) for different stimulus types (correct performance on conflict trials is undefined). Plotted as in (b). $*$: $p < 10^{-6}$ (paired t-test), n.s.: $p > 0.05$* ***(f,g)*** *Same data and fits as c,d, replotted as the odds of choosing right vs. left (in log coordinates, Y-axis), as a function of visual contrast raised to the power γ. In this space, the predictions of the model are straight lines.* ***(h)*** *Log_2-likelihood ratio for the additive, visual-only ($g(A) = 0$) and auditory-only ($f(V) = 0$) models versus a full model where each combination of visual and auditory stimuli is allowed its own behavioural response (each assessed by 5-fold cross-validation relative to a bias-only model). Triangles and diamonds indicate left and right example mice from (c), squares indicate combined fit across 17 mice. Visual-only and auditory-only models are significantly worse than the full model ($p < 10^{-6}$; 17 mice, paired t-test). There is no significant difference between the additive and full models ($p > 0.05$).*

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This additive behavioural model provided excellent fits to the multisensory decisions of all the mice (Fig. 1c-g, Extended Data Fig. 2a-o). The additive model was able to fit the choices of individual mice (Fig. 1c) as well as the mean choice averaged across mice (Fig. 1d). A particularly clear view of the additive model can be obtained by replotting these data in terms of log odds of rightward vs. leftward choices (as in $Eqn. 1$) vs. linearized visual contrast (contrast raised by the exponent γ, $Eqn. 3$). In this representation, the responses to unisensory visual stimuli fall on a line, and the effect of adding auditory cues is to shift this line additively (Fig. 1f,g, Extended Data Fig. 2a-o). The intercept of the line is determined by the bias b, its slope by the visual sensitivity v, and the additive offset by the auditory sensitivity a.

The additive behavioural model performed as well as more complex models allowing for non-additive interactions, indicating that mice integrated auditory and visual cues additively (Fig. 1h). We tested the additive model by comparing its cross-validated likelihood to that of a full model with a parameter for each stimulus combination (Fig. 1h), which did not show higher performance despite having substantially more parameters. However, the full model out-performed both visual-only ($g(A) = 0$) and auditory-only ($f(V) = 0$) models (Fig. 1h). Fitting the model parameters from unisensory trials alone was also sufficient to predict behaviour on multisensory trials (Extended Data Fig. 2p). The additive model excludes non-additive interactions between the senses, indicating that mice did not default to being purely auditory as has been reported[14]. Indeed, if mice defaulted to auditory cues in conflict trials, the red and blue curves would be horizontal when visual contrast was on the left and right.

Optogenetic inactivation identifies roles of sensory and frontal cortical areas

To determine which regions of dorsal cortex were necessary for the performance of our audiovisual task, we used laser-scanning photo-inactivation (Fig. 2a). We optogenetically inactivated 52 cortical sites in mice expressing ChR2 in Parvalbumin interneurons, by transcranial laser illumination[55–58] (3 mW; 462 nm; 1.5 s duration following stimulus onset; Fig. 2a). We combined results across mice and hemispheres as inactivation effects were qualitatively consistent and symmetric (Extended Data Fig. 3a-b). Control measurements established that mouse choices were not affected when we targeted inactivations to regions outside the brain (Extended Data Fig. 3c).

Inactivation of primary sensory cortices only impaired decisions based on the corresponding modality, whereas inactivation of frontal area MOs impaired decisions based on either modality (Fig. 2b-e). When a unisensory visual stimulus was presented, the fraction of incorrect choices was increased by inactivation of the contralateral visual areas (VIS) (Fig. 2b, consistent with prior work from visual detection tasks[49,55,59]). Conversely, when a unisensory auditory stimulus was presented, the fraction of incorrect choices was increased by inactivation of contralateral areas proximal to auditory cortex (AUD), consistent with evidence that auditory cortex is required for auditory localization[60] (Fig. 2c). For both types of unisensory stimulus – visual or auditory – the fraction of incorrect choices was also increased by inactivating contralateral secondary motor area (MOs). These roles of sensory areas VIS and AUD and of frontal area MOs extended to multisensory trials: inactivation of sensory cortices increased choices away from contralateral stimuli of the appropriate modality, while inactivation of MOs increased choices away from contralateral stimuli of either modality (Fig. 2d-e). We did not observe multisensory effects of inactivation in any additional region: MOs was the only area in dorsal cortex that the mouse required to respond to stimuli of both modalities.

The results of these inactivations were well predicted by the additive model, implicating MOs in the additive integration of visual and auditory signals (Fig. 2f-i). Inactivating primary visual cortex (VISp) in one hemisphere strongly reduced the performance of the mouse for contralateral visual stimuli, regardless of auditory stimuli (Fig. 2f-g). The model accounted for this via a decrease in the sensitivity for contralateral visual stimuli (v_c; Fig. 2i). Inactivating AUD had an analogous effect on contralateral auditory sensitivity (Extended Data Fig. 3d), although this effect was weaker, presumably due to the curvature of the cortex causing weaker laser illumination. Finally, inactivating MOs had a qualitatively different effect, increasing the probability of ipsilateral choices in all stimulus conditions (Fig. 2h). The model accounted for this effect via a concomitant reduction in both visual and auditory sensitivity together with an increase in bias (Fig. 2i). The model thus can explain the asymmetric effects of MOs inactivation[54].

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

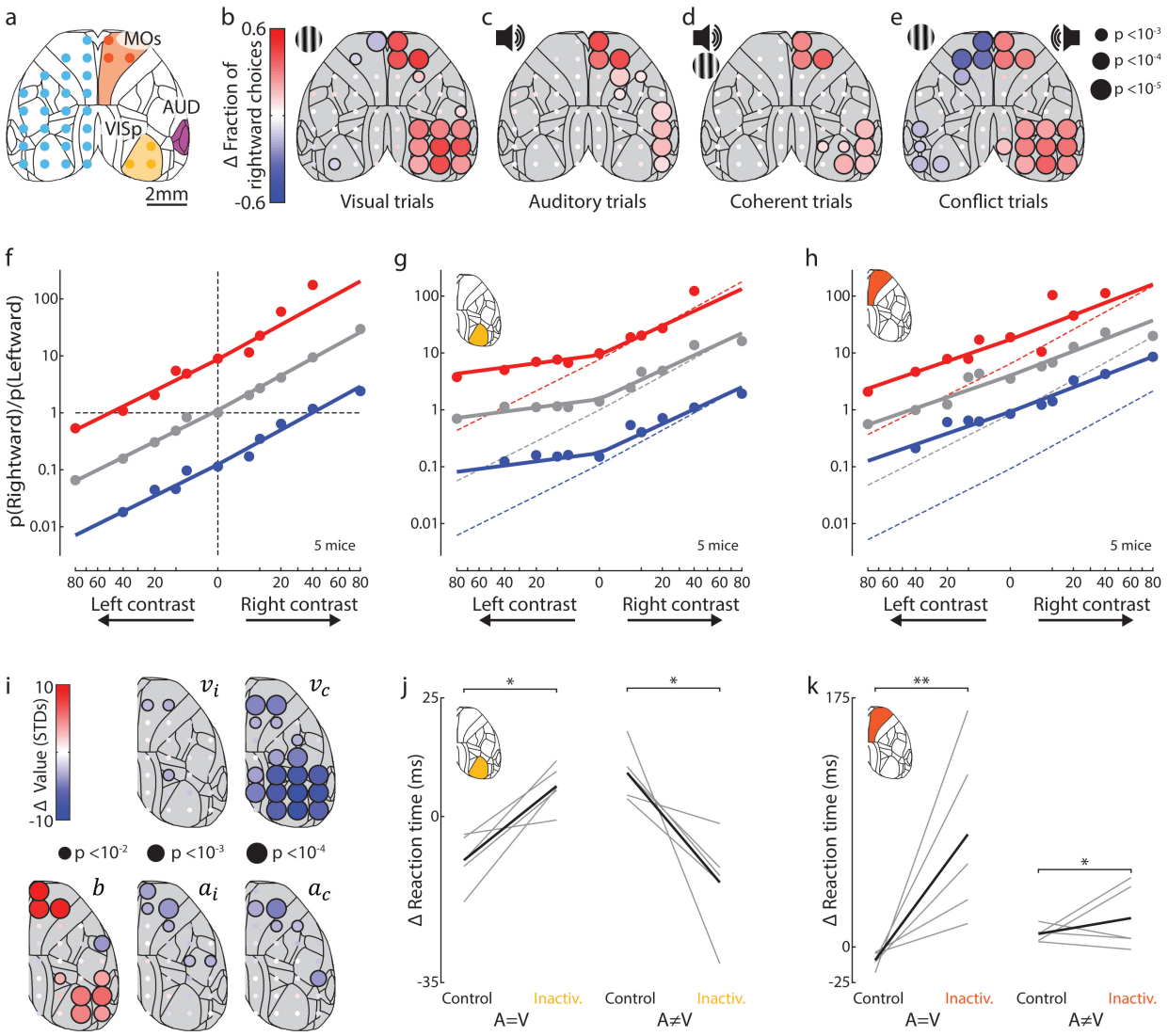


Fig. 2 Optogenetic inactivation identifies roles of sensory and frontal cortical areas

(a) *Schematic of inactivation sites. On ~75% of trials, a blue laser randomly illuminated regions centred on one of 26 locations in each hemisphere (blue dots, shown only in one hemisphere for graphical purposes), from stimulus onset (fixed duration of 1.5 s). Yellow and orange dots indicate sites centred in the primary visual area (VISp) and in the secondary motor area (MOs). Sites that inactivated the dorsal aspect of auditory cortex (purple) also inactivated lateral visual cortex.* ***(b)*** *Change in the fraction of rightward choices for 52 laser sites for unisensory left visual stimulus trials. Dot colour indicates change in the fraction of rightward choices; dot size represents statistical significance (5 mice, permutation test, see Methods). Data for right stimulus trials were also included in the average after reflecting the maps (see Extended Data Fig. 3a for both individually).* ***(c)*** *As in (b), but for unisensory auditory trials.* ***(d)*** *As in (b), but for coherent multisensory trials.* ***(e)*** *As in (b), but for conflict multisensory trials.* ***(f)*** *Fit of the additive behavioural model to non-inactivation trials combined across 5 mice, plotted as in Fig. 1g.* ***(g)*** *As in (f), but for trials when right VISp was inactivated, combined across the three sites shown in panel (a). Dashed lines indicate fits from (f). Trials with inactivation of left VISp were also included in the average after reflecting the maps (5 mice, 6497 trials).* ***(h)*** *As in (g), but for trials when MOs was inactivated (5 mice, 5612 trials).* ***(i)*** *As in (b-e), but dot colour indicates the change in parameters of the additive behavioural model. R and L (for stimuli/choices on the right and left) are replaced with i and c (for stimuli/choices that are ipsilateral and contralateral to the inactivated side).* ***(j)*** *Change in multisensory reaction times when VISp was inactivated contralateral to the visual stimulus. Grey and black lines indicate individual mice (n = 5) and the mean across mice. Median reaction times for each stimulus type are shown relative to the mean across the stimulus types in the control condition. Inactivation significantly increased reaction time on coherent trials and decreased it on conflict trials. ∗: $p < 0.001$ (linear mixed effects model).* ***(k)*** *As in (j), but for inactivations of MOs contralateral to the visual stimulus (note different y-axis scale to j). Inactivation significantly increased reaction time on both coherent and conflict trials. ∗: $p < 0.05$, ∗∗: $p < 0.001$ (linear mixed effects model).*

The hypothesis that MOs integrates multisensory information was further reinforced by the effects of inactivations on reaction times (Fig. 2j,k). As we have seen, reaction times were longer during multisensory conflict trials than either coherent or unisensory auditory trials (Fig. 1b). Inactivating VISp increased reaction times on coherent trials but decreased them for conflict trials (Fig. 2j). This result is consistent with the mouse perceiving a multisensory stimulus as unisensory auditory, effectively turning both conflict and coherent trials into unisensory auditory trials. In contrast, inactivation of MOs increased reaction times on both conflict and coherent trials (Fig. 2k). This result is consistent with a role for MOs in both unisensory and multisensory processing.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

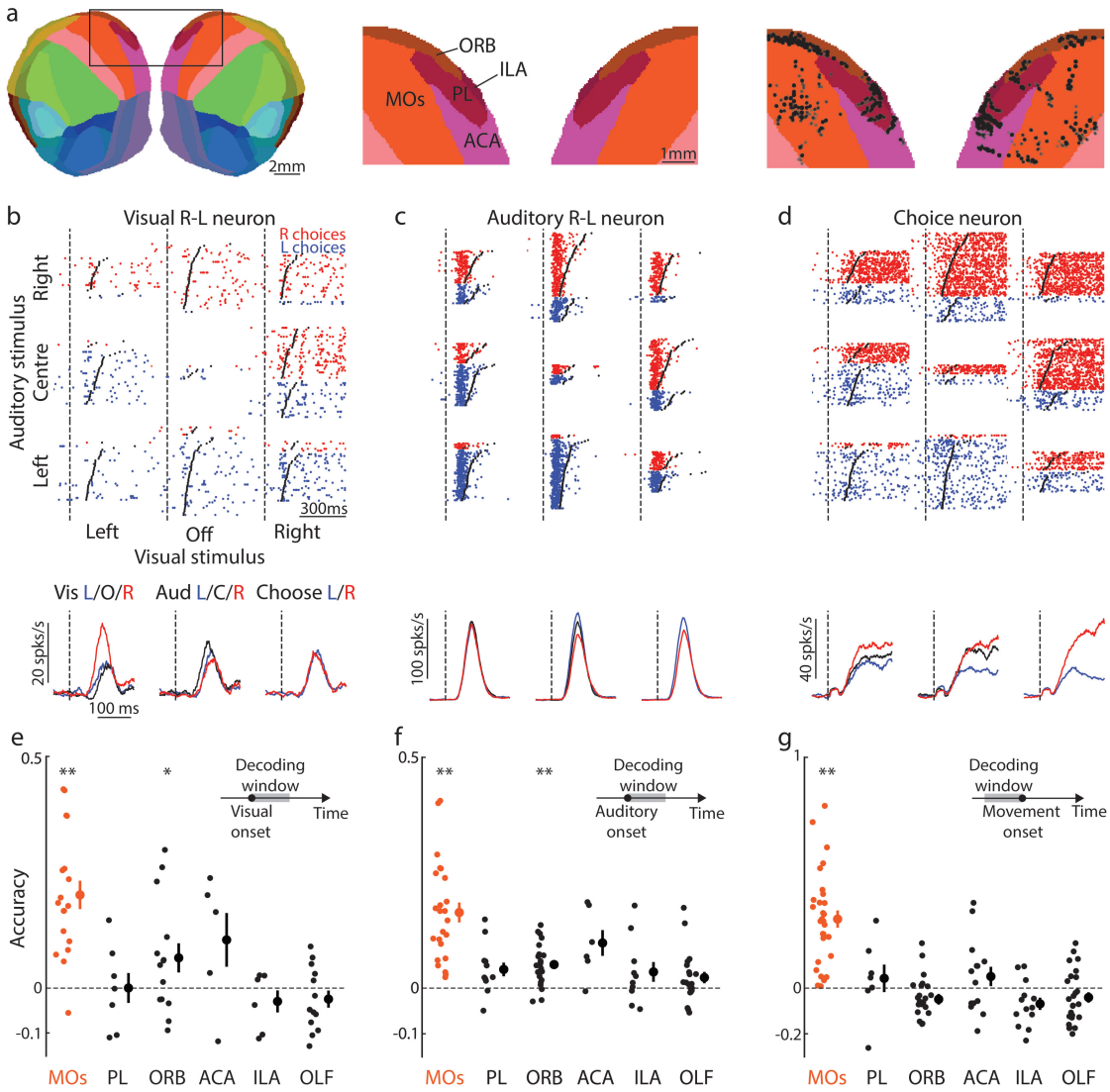


Fig. 3 Neurons in area MOs encode stimuli and predict behaviour

(a) *Recording locations for cells (black dots, right), overlayed on a flattened cortical map (using the Allen CCF[87]), showing locations in secondary motor (MOs), orbitofrontal (ORB), anterior cingulate (ACA), prelimbic (PL) and infralimbic (ILA) areas.* ***(b)*** *Top: Raw spike rasters, separated by trial condition, from a neuron sensitive to visual spatial location. Red/blue rasters indicate trials where the mouse made a rightward/leftward choice. Dashed line and black points represent stimulus onset and movement initiation. Bottom: PSTH of the neural response, averaged across different visual (left) auditory (centre) or choice (right) conditions. Trials are not balanced: choice and stimulus location can be correlated.* ***(c)*** *As in (b), for a neuron that is sensitive to auditory spatial location.* ***(d)*** *As in (b), for a neuron that is sensitive to the animal's choice.* ***(e)*** *Cross-validated accuracy (relative to a bias model, see Methods) of an SVM decoder trained to predict visual stimulus location from population spiking activity time-averaged over a window 0 ms to 300 ms after stimulus onset. Accuracies 0 and 1 represent chance and optimal performance. Each point represents the decoding accuracy from neurons in one of the regions labelled in (a), or olfactory areas (OLF), from a single experimental session. For each point, 30 neurons were randomly selected to equalize population size across sessions and brain regions. $**$: $p < 0.0001$, $*$: $p < 0.01$ ($\geq$ 5 sessions from 2-5 mice for each region, t-test).* ***(f)*** *As in (e), but for decoding of auditory stimulus location ($\geq$ 6 sessions, 3-6 mice).* ***(g)*** *As in (e), but for decoding choices based on spiking activity time-averaged over a window 0 ms to 130 ms preceding movement onset ($\geq$ 7 sessions, 3-6 mice).*

Neurons in area MOs encode stimuli and predict behaviour

To investigate the neural code that MOs might use to integrate multisensory information, we recorded acutely with Neuropixels probes during behaviour (Fig. 3a-d). We recorded 10,548 neurons from frontal cortex across 88 probe insertions (56 sessions) from 6 mice (Fig. 3a, Extended Data Fig. 4a-b). Of these neurons, 2581 were in area MOs, and the rest were in neighbouring areas: orbitofrontal (ORB), anterior cingulate (ACA), prelimbic (PL) and infralimbic (ILA) and in the olfactory areas (OLF). We observed a variety of neural responses, including neurons that were sensitive to visual and auditory location (Fig. 3b-c), and to the animal's upcoming choice (Fig. 3d).

MOs neurons encoded task information more accurately than neurons in other frontal regions (Fig. 3e-g). To identify which regions encoded task information, we trained separate linear support vector machine (SVM) decoders to predict stimulus location or upcoming choice from neural activity. MOs had the strongest encoding of auditory and visual

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

stimulus location (Fig. 3e-f) and was the only region able to predict the animal's upcoming choice before movement onset (Fig. 3g). These observations were robust to the correlations between stimuli and choices: even when controlling for this correlation, MOs still had the highest percentage of neurons with significant stimulus-location or pre-movement choice coding (Extended Data Fig. 4e-f). Once movements were underway, however, we could decode their direction from all regions, consistent with prior observations that ongoing movements are encoded throughout the brain[61,62] (Extended Data Fig. 4c).

Neurons in area MOs additively encode task variables

Given the additive effects of visual and auditory signals on behaviour, we hypothesized that they may be combined additively in MOs neural activity. To test this hypothesis, we used an ANOVA-style decomposition to analyse responses to combined audiovisual stimuli during behaviour. For simplicity, we focused on trials of a single visual contrast, so that we could define binary variables $a_i, v_i, c_i = \pm 1$ to encode the laterality (left vs. right) of auditory stimuli, visual stimuli, and choices. We can then decompose $\boldsymbol{F}_i(t)$, the firing rate on trial i at time t after stimulus onset, as the sum of 6 temporal kernels:

$$\boldsymbol{F}_i(t) = \boldsymbol{B}(t) + a_i\boldsymbol{A}(t) + v_i\boldsymbol{V}(t) + a_i v_i\boldsymbol{N}(t) + \boldsymbol{M}(t-\tau_i) + c_i\boldsymbol{D}(t-\tau_i) \qquad (5)$$

Here, $\boldsymbol{B}$ is the overall mean stimulus response, $\boldsymbol{A}$ and $\boldsymbol{V}$ are the additive main effects of auditory and visual stimulus location, and $\boldsymbol{N}$ is a non-additive interaction between them. To account for the effects of movement, $\boldsymbol{M}$ is a kernel for the mean effect of movement (regardless of direction, and relative to τ_i, the time of movement onset on trial i) and $\boldsymbol{D}$ is the differential effect of movement direction (right minus left). To test for additivity, we compared performance of this full model against a restricted additive model, constrained by setting the non-additive interaction term to 0 ($\boldsymbol{N} = 0$), and we used cross-validation to compare the two models.

The additive neural model performed better than more complex alternatives, suggesting that MOs uses an additive code to integrate visual and auditory signals for stimulus location (Fig. 4a-f). The restricted additive model performed better than the full model with interactions between visual and auditory stimuli (Fig. 4a-c), and better than an alternative full model with interactions between stimuli and movement (Extended Data Fig. 5d). Similar results were seen during passive presentation of task stimuli, when sensory responses could not be confounded by activity related to movement (Fig. 4d-f, Extended Data Fig. 5h). MOs thus encodes the location of visual and auditory stimuli in an additive fashion.

MOs neurons provided a mixed representation of visual and auditory stimulus locations, but encoded the two modalities with different time courses (Fig. 4g-h). Consistent with reports of mixed multisensory selectivity in parietal cortex of rat[24] and primate[5], the auditory and visual stimulus preferences of cells in MOs were neither correlated nor lateralized: cells in either hemisphere could represent the location of auditory or visual stimuli with a preference for either location, and could represent the direction of the subsequent movement with a preference for either direction (Fig. 4g, Extended Data Fig. 5a-c,f-g). Neurons that responded to one modality, however, also tended to respond to the other, as evidenced by a weak correlation in the absolute sizes of the auditory and visual kernels: (Extended Data Fig. 5e). Nevertheless, representations of auditory and visual stimuli had different time courses: the time at which neurons could distinguish stimulus presence and location was sooner for auditory than for visual stimuli (Fig. 4h), consistent with earlier behavioural reactions to auditory stimuli (Fig. 1b). Furthermore, information about auditory onset was stronger, and available significantly earlier than information about visual onset or the location of either stimulus (Fig. 4h, Extended Data Fig. 5i). This could explain why past work using a detection task, rather than a localization task, observed strong auditory dominance in multisensory conflict trials[14].

An accumulator model trained on passive MOs activity reproduced decisions

Having established that MOs satisfies key criteria for a region integrating visual and auditory inputs, we finally asked whether an accumulator model trained on its neuronal responses could reproduce behavioural choices (Fig. 5a,b). The previous analyses established that MOs neurons linearly sum information about stimulus location, with timing consistent with the observed behavioural responses. To test whether this activity could underpin multisensory behaviour, we asked whether an accumulator model taking MOs population activity as its input would behave similarly

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

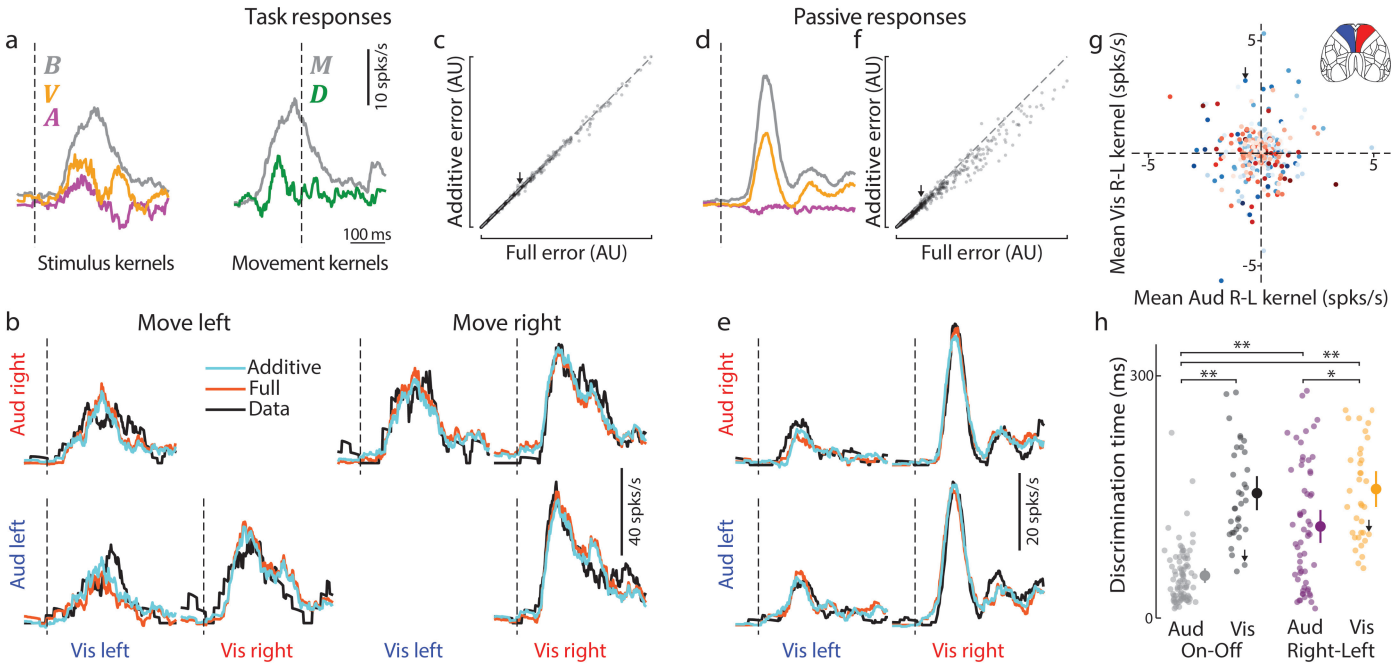


Fig. 4 Neurons in area MOs additively encode task variables

(a) *Example kernels from fitting the additive neural model to an example neuron (setting the non-additive kernel ($\boldsymbol{N}$) $= 0$). Dashed lines indicate stimulus onset for stimulus kernels and movement onset for movement kernels.* ***(b)*** *Cross-validated model fits for the example neuron from (a) to neural activity averaged over each combination of stimuli and movement direction. Trials where stimuli were coherent but the mouse responded incorrectly were excluded due to low trial numbers. Cyan and orange lines show predictions of the additive ($\boldsymbol{N} = 0$) and full models, black line shows test-set average responses. Dashed lines indicate stimulus onset.* ***(c)*** *Prediction error (see Methods) across all neurons for the additive and full models. Arrow indicates position of the example cell in (a-b). The additive model has smaller prediction error (p = 0.037, linear mixed effects model, 2183 cells, 5 mice). The largest 1% of errors were excluded from plot for visualization, but not from statistical analysis.* ***(d)*** *As in (a), but for neural activity during passive stimulus presentation, using only sensory kernels.* ***(e)*** *As in (b), but fitting sensory kernels to all audiovisual combinations presented in passive conditions.* ***(f)*** *As in (c), but for passive stimulus conditions. The additive model has smaller prediction error. $p < 10^{-10}$ (2509 cells, 5 mice, linear mixed-effects model).* ***(g)*** *Encoding of visual stimulus preference (time-averaged amplitude of kernel, $\boldsymbol{V}$) versus encoding of auditory stimulus preference (time-averaged amplitude of kernel, $\boldsymbol{A}$) for each cell. There was no correlation between auditory and visual directional preference. $p > 0.05$ (2509 cells, Pearson correlation test). Red/blue indicates whether the cell was recorded in the right/left hemisphere and saturation indicates fraction of variance explained by the sensory kernels.* ***(h)*** *Discrimination time (see Methods) relative to stimulus onset during passive conditions. Auditory Right-Left neurons (sensitive to auditory location, magenta, n = 59), differentiated earlier than Visual Right-Left neurons (gold, n = 36). Auditory On-Off neurons (sensitive to auditory presence, but not necessarily location, grey, n = 82) discriminated earlier than all other neuron classes, including Visual On-Off (n = 36, black). Points are individual neurons and bars indicate standard error. $*$: $p < 0.01$, $**$: $p < 10^{-7}$ (Mann–Whitney U test).*

to the mice (Fig. 5a,b). To avoid the confound of movement encoding in MOs we used passive stimulus responses, generating a population representation by selecting (from all recordings) MOs neurons encoding the location of at least one of the sensory stimuli, and using their passive stimulus responses to generate surrogate population spike trains $\boldsymbol{x}(t)$. These spike trains were fed into an accumulator model (akin to a drift diffusion model[1,63,64]): they were linearly integrated over time to produce a scalar decision variable $d(t)$:

$$d(t) = d(t-1) + \boldsymbol{x}(t) \cdot \boldsymbol{w} \qquad (6)$$

The model was judged to make a choice when $d(t)$ crossed a decision boundary. The weight vector $\boldsymbol{w}$ was found as that producing the fastest and most accurate choices possible given the MOs representation $\boldsymbol{x}(t)$; the actual behaviour of the mouse was not used to train the model. The accumulator model applied to MOs activity closely mimicked mouse behaviour and predicted the effects of cortical inactivation (Fig. 5c-e). As observed with mice, the model integrated multisensory stimuli additively (Fig. 5c; cf. Fig. 1g). Selectively suppressing left-preferring visual neurons in our model reproduced the effects of inactivating right VISp (Fig. 5d; cf. Fig. 2g). Strikingly, the model also accounted for the different behavioural reaction times for different stimulus types: faster in auditory and coherent trials, and slower in visual and conflict trials (Fig. 5e; cf. Fig. 1b). These results demonstrate that the additive coding and relative sensory timing observed in MOs neurons have the form required to underpin the subjects' behaviour, further implicating MOs as the cortical site of audiovisual integration.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

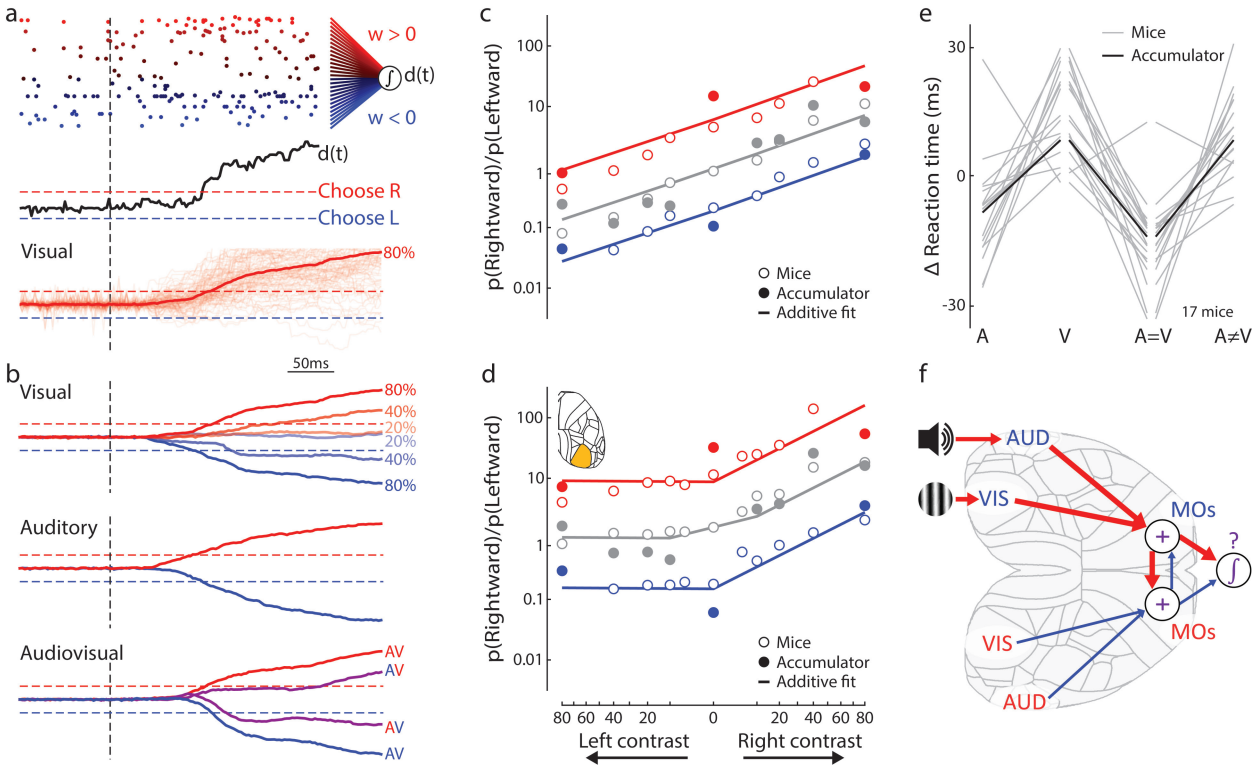


Fig. 5 An accumulator model trained on passive MOs activity reproduced decisions

(a) *An example trial from the accumulator model. Top: population spike train rasters for a single trial, coloured according to the fitted weight for that neuron, with red and blue indicating that spiking will push the decision variable, d_t, to the rightward or leftward decision boundary. Vertical dashed line represents stimulus onset. Middle: The evolution of the decision variable over the course of a single trial. Red/blue dashed lines indicate the rightward/leftward decision boundaries. Bottom: Individual examples of the decision variable (thin) and the mean result (thick) for unisensory visual trials where rightward contrast was 80%.* ***(b)*** *The mean trajectory of the decision variable on visual-only (top), auditory-only (middle), and multisensory (bottom) stimulus conditions.* ***(c)*** *Mean behaviour of the accumulator model (solid circles), together with mouse performance (open circles, n = 17, plotted as in Fig. 1g). Solid lines represent the fit of the additive behavioural model to the accumulator model output.* ***(d)*** *As in (c), but for an accumulator model with the weights of visual-left-preferring cells reduced by 50% to simulate inactivation of right VISp (solid circles), plotted with the mean behaviour from VISp-inactivated mice (5 mice, open circles, cf. Fig. 2g).* ***(e)*** *Median reaction times for different stimulus types, relative to mean over all conditions, for the accumulator model (black lines) and mice (grey, n=17, cf. Fig. 1b).* ***(f)*** *Diagram of hypothesised audiovisual integration pathway through cortex.*

Discussion

We found that mice integrate auditory and visual cues to spatial location using an additive law, and that frontal area MOs is the key cortical region for this additive integration. Indeed, we found that inactivation of MOs impaired audiovisual decisions, that MOs neurons combine visual and auditory location signals additively, and that an accumulator model applied to MOs activity predicts the direction and timing of behavioural responses.

Our result that mice integrate multisensory signals additively extends the evidence for optimal integration observed in other species[2–6,8,9], but might appear to contradict previous observations in mice, where prior work using an audiovisual detection task had suggested that mice were auditory dominant rather than Bayes-optimal[14]. We hypothesize that this difference arises from differences between the tasks. Our task required audiovisual localization, and we found that neural signals representing stimulus locations were combined additively in MOs, with subtle differences in the encoding of auditory and visual stimulus location able to explain the mice's earlier reactions to auditory stimuli. However, we also saw that neural signals encoding auditory onset were stronger and substantially earlier than neural signals encoding either visual onset or stimulus location from either modality. This suggests that in a detection task[14], strong and early auditory onset signals might be able to dominate behaviour. Under this hypothesis, there is no contradiction between our results and the previous ones: mice might integrate audiovisual signals additively when tasked with localizing a source but be dominated by auditory cues when tasked with detecting the source's presence.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Inactivation and recordings identified frontal area MOs as the main audiovisual integration site in dorsal cortex, in contrast to studies that highlighted roles of parietal or sensory cortices. Scanning optogenetic inactivation indicated that MOs was the only region of dorsal cortex required to respond to both modalities, and electrode recordings indicated that MOs was the only region of frontal cortex to encode information about both modalities as well as choices. This might appear to contradict previous work implicating parietal cortex in multisensory integration[3,5,6,14,24–32], or showing multisensory activity in primary sensory cortices[36–45]. However, our finding agrees with evidence that parietal cortex can reflect multisensory activity without being causally involved in a task[24,33,46,47]. Our data further imply that the causal role of visual and auditory cortices in this task is restricted to processing stimuli of the corresponding modality and relaying the signals to other sites where the two information streams are integrated[28,33,35].

MOs neurons additively integrate the locations of visual and auditory stimuli, producing a population representation that is formatted suitably for guiding a downstream decision circuit. Secondary motor cortex has been linked to multiple functions in rodents[65], including flexible sensory-motor mapping[66,67], perceptual decision-making[55,56,68–71], value-based action selection[72], and the exploration-exploitation trade-off in both visual and auditory behaviours[54]; furthermore, homologous regions of frontal cortex can encode multisensory information in primates[34]. We found that MOs additively integrated the locations of visual and auditory stimuli (even under passive conditions), as required to compute a probabilistic representation of the estimated stimulus location probabilities[73]. The resulting population code, fed into a simulated decision circuit, produced choices and timing similar to those of the mice. We hypothesize that downstream of the MOs sensory representation lies a circuit that accumulates and thresholds this activity to select an appropriate action; based on previous studies in a unisensory task[74] we suspect this circuit may include neurons of basal ganglia and midbrain, besides neurons in MOs itself.

The format of this MOs code has some apparently paradoxical features, but these would not prevent its efficient use by a downstream accumulator. First, a neuron's preference for visual stimulus location showed no apparent relation to its preference for auditory stimulus location, consistent with previous reports of multisensory neural populations in primates[5] and rats[24]. This suggests that most neurons in MOs form a representation of "mixed selectivity", combining information from different modalities apparently at random. Such a code might result from random network connectivity, which can be surprisingly computationally powerful[75,76], as the resulting high-dimensional mixed representations may allow downstream circuits to quickly learn to extract behaviourally relevant feature combinations. In the present case, neurons encoding incoherent stimulus locations would not prevent a downstream decision circuit from learning to respond correctly; they could be ignored in the current task, but their presence would provide flexibility should task demands later change. Second, although an approximately equal number of neurons in each hemisphere preferred left and right stimuli of either modality, inactivation of MOs caused a lateralized effect on behaviour. This apparent contradiction could be resolved if the downstream decision circuit weights MOs neurons with a hemispheric bias, or if particular projection neurons showed lateral bias as has been seen in a somatosensory task[77].

In summary, our data suggest that audiovisual integration in our task has an extremely simple form, with processing of unisensory stimuli in sensory cortical areas followed by simple additive operations in frontal area MOs (Fig. 5f). Both input streams converge on MOs, whose neurons combine them linearly and randomly. This simple combination rule suffices to generate a representation of the log odds of the stimulus location given the sensory input, which can then be read out by downstream circuits that accumulate spikes over time to form a decision. While our current data cannot rule out parallel subcortical pathways, it appears unlikely that other cortical regions are involved in multisensory processing: the causal role of visual and auditory cortices, which both project to MOs[54,66,78], might simply reflect their supplying MOs with spatially-selective inputs. The circuits that convert this MOs representation of stimulus odds into a thresholded decision remain to be established but may include basal ganglia and midbrain, as well as neurons within MOs itself.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

We thank Michael Krumin, Peter Zatka-Hass, Andrew J Peters, Max Hunter, James Chadwick, Paul Jaohnson, and Ian Macartney for assistance with the experimental setup; Charu Reddy, Laura Funnell, Dylan Rich, Siddharth Kackar and Hamish Forrest for help with mouse husbandry and training; Charu Reddy for surgical assistance and optimization; Laura Funnell, Rakesh Raghupathy, and David Orme for histology processing; Samuel Picard, Célian Bimbard, and Maxwell Shinn for reading earlier versions of the manuscript. This work was supported by the Wellcome Trust (Sir Henry Wellcome Postdoctoral Fellowship 531073 to PC, and grants 205093 and 204915 to MC and KDH), the European Research Council (694401 to KDH), and the European Union's Horizon 2020 research and innovation programme (Marie Sklodowska-Curie fellowship 531949 to PC). TS is supported by the Sainsbury Wellcome Centre PhD program. MC holds the GlaxoSmithKline/Fight for Sight Chair in Visual Neuroscience.

Author contributions

PC, MC, and KDH conceived of and designed the study. PC and MJW collected the behavioural data and developed the behavioural task. PC collected the neural data. PC and TS analysed the behavioural and neural data. PC and KDH wrote the manuscript, TS and MC provided extensive edits. Correspondence and material requests should be directed to Philip Coen, p.coen@ucl.ac.uk.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Experimental procedures were conducted according to the UK Animals Scientific Procedures Act (1986) and under personal and project licenses released by the Home Office following appropriate ethics review.

Terminology

Here, we define some terms used throughout the methods and manuscript. A "stimulus condition" refers to a particular combination of auditory and visual stimuli; for example, a visual stimulus of 40% contrast on the left and an auditory stimulus presented on the right. A "stimulus type" refers to a category that may comprise several stimulus conditions. We define five different stimulus types: unisensory auditory, unisensory visual, coherent, conflict, and neutral. "Unisensory auditory" trials are when an auditory stimulus is presented on the left or right, and visual contrast is zero (grey screen). "Unisensory visual" trials are when a stimulus of any contrast greater than zero is presented on the left or right, and the auditory stimulus is presented in the centre (during behaviour) or is absent (during passive conditions). "Coherent" trials are when a visual stimulus with non-zero contrast is presented on the same side as an auditory stimulus. "Conflict" trials are when a visual stimulus with non-zero contrast is presented on a different side from an auditory stimulus. "Neutral" trials are when the visual contrast is zero and the auditory stimulus is presented in the centre. We refer to a single experimental recording (whether purely behaviour, or combined with optogenetic inactivation or electrophysiology) as a "session." Sessions can vary in duration and number of trials.

Mice

Experiments were performed on 3 male and 15 female mice, aged between 9 and 21 weeks at time of surgery. For all experiments, we used transgenic mice expressing ChR2 in Parvalbumin-positive inhibitory interneurons (Ai32 [Jax #012569, RRID:IMSR_JAX: 012569] x PV-Cre [Jax #008069, RRID:IMSR_JAX: 008069]). 17 mice contribute to behavioural data (Fig. 1), 5 mice contribute to optogenetic inactivation data (Fig. 2), and 6 mice contribute to electrophysiological recordings (Fig. 3-5). Behavioural data (Fig. 1) comprised both sessions without any optogenetic inactivation and non-inactivation trials within optogenetic experiments.

Surgery

A brief (around 1 h) initial surgery was performed under isoflurane (1–3% in O2) anaesthesia to implant a steel headplate (approximately 25 × 3 × 0.5 mm, 1 g) and, in most cases, a 3D-printed recording chamber. The chamber comprised two pieces of opaque polylactic acid which combined to expose an area approximately 4 mm anterior to 5 mm posterior to bregma, and 5 mm left to 5 mm right, narrowing near the eyes. The implantation method largely followed Guo et al.[56] and has been previously described[79]. In brief, the dorsal surface of the skull was cleared of skin and periosteum. The lower part of the chamber was attached to the skull with cyanoacrylate (VetBond; World Precision Instruments) and the gaps between chamber skull were filled with L-type radiopaque polymer (Super-Bond C&B, Sun Medical). A thin layer of cyanoacrylate was applied to the skull inside the cone and allowed to dry. Thin layers of UV-curing optical glue (Norland Optical Adhesives #81, Norland Products) were applied inside the cone and cured until the exposed skull was covered. The head plate was attached to the skull over the interparietal bone with Super-Bond polymer. The upper part of the cone was then affixed to the headplate and lower cone with a further application of polymer. After recovery, mice were treated with carprofen for three days, then acclimated to handling and head-fixation before training.

Audiovisual behavioural task

The two-alternative unforced choice task design was an extension of a previously described visual task[49]. It was programmed in Signals, part of the Rigbox MATLAB package[80]. Mice sat on a plastic apparatus with their forepaws on a rigid, rubber Lego wheel affixed to a rotary encoder (Kubler 05.2400.1122.0360). A plastic tube for delivery of water rewards was placed near the subject's mouth.

Visual stimuli were presented using three computer screens (Adafruit, LP097QX1), arranged at right angles to cover ±135° azimuth and ±45° elevation, where 0° is directly in front of the subject. Each screen was roughly 11 cm from the mouse's eyes at its nearest point and refreshed at 60 Hz. Intensity values were linearized[49] with a photodiode (PDA25K2, Thor labs). The screens were fitted with Fresnel lenses (Wuxi Bohai Optics, BHPA220-2-5) to ameliorate reductions in luminance and contrast at larger viewing angles, and these lenses were coated with scattering window

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

film ('frostbite', The Window Film Company) to reduce reflections. Visual stimuli were flashing vertical Gabors presented with a 9° Gaussian window, spatial frequency 1/15 cycles per degree, vertical position 0° (i.e. level with the mouse) and phase randomly selected on each trial. Stimuli flashed at a constant rate of 8Hz, with each presentation lasting for ~50 ms (with some jitter due to screen refresh times).

Auditory stimuli were presented using an array of 7 speakers (102-1299-ND, Digikey), arranged below the screens at 30° azimuthal intervals from -90° to +90° (where -90°/+90° is directly to the left/right of the subject). Speakers were driven with an internal sound card (STRIX SOAR, ASUS) and custom 7-channel amplifier (http://maxhunter.me/portfolio/7champ/). The frequency response of each speaker was individually estimated in situ with white noise playback recorded with a calibrated microphone (GRAS 40BF 1/4" Ext. Polarized Free-field Microphone). For each speaker, a compensating filter was generated to flatten the frequency response using the Signal Processing Toolbox in MATLAB. Throughout all sessions, we presented white noise at ~50 dbSPL to equalize background noise between different training and experimental rigs.

Auditory stimuli were 50 ms pulses of filtered pink noise (8-16kHz, 75-80 dbSPL), with 16ms sinusoidal onset/offset ramps. To ensure mice did not entrain to any residual difference in the frequency response of the speakers, auditory stimuli were further modulated on each trial by a filter selected randomly from 100 pre-generated options, which randomly amplified and suppressed different frequency components within the 8-16kHz range. As with visual stimuli, sound pulses were presented at a rate of 8Hz. On multisensory trials, the modulation of visual and auditory stimuli were synchronized, but software limitations and hardware jitter resulted in visual stimuli marginally preceding auditory stimuli by 10 $\pm$ 12 ms (mean $\pm$ s.d.).

A trial was initiated after the subject held the wheel still for a short quiescent period (duration uniformly distributed between 0.1 and 0.25 s on each trial; Fig. 1a). Mice were randomly presented with different combinations of visual and auditory stimuli (Extended Data Fig. 1a). Visual stimuli varied in azimuthal position (-60° or +60°) and contrast (0%, 10%, 20%, 40%, and 80%, and also 6% in a subset of mice). On unisensory auditory trials, visual contrast was zero (grey screen). Auditory stimuli varied only in azimuthal position: -60°, 0°, or +60°; on unisensory visual trials, auditory stimuli were positioned at 0°. A small number of "neutral trials" had zero visual contrast, and an auditory stimulus at 0°. The ratio of unisensory visual/unisensory auditory/multisensory coherent/multisensory conflict/neutral trials varied between sessions but was ~ 10/10/5/5/1, and stimulus side was selected randomly on each trial.

After stimulus onset there was a 0.5 s open-loop period, during which the subject could turn the wheel without penalty, but stimuli were locked in place and rewards could not be earned. The mice nevertheless typically responded during this open-loop period (Extended Data Fig. 1f). At the end of the open-loop period, an auditory Go cue was delivered through all speakers (10 kHz pure tone for 0.1 s) and a closed-loop period began in which the stimulus position (visual, auditory, or both) became coupled to movements of the wheel. Wheel turns in which the top surface of the wheel was moved to the subject's right led to rightward movements of stimuli on the speaker array and/or screen, that is, a stimulus on the subject's left moved towards the central screen. For visual or auditory stimuli, the position updated at the screen refresh rate (60Hz) or the rate of stimulus presentation (8Hz). In trials, where auditory stimuli were presented at 0°, the auditory stimulus did not move throughout the trial. A left or right turn was registered when the wheel was turned by an amount sufficient to move the stimulus by 60° in either azimuthal direction (~30° of wheel rotation, although this varied across mice/sessions); if this had not occurred within 1 s of the auditory Go cue, the trial was recorded as a "timeout." On unisensory visual, unisensory auditory, and multisensory coherent trials, the subject was rewarded for moving the stimulus to the centre. If these trials ended with an incorrect choice, or a timeout, then the same stimulus conditions were repeated up to a maximum of 9 times. In neutral and conflicting multisensory trials, left and right turns were rewarded with 50% probability (Extended Data Fig. 1a), and trials were only repeated in the event of a timeout, not an unrewarded choice. An incorrect choice or timeout resulted in an extra 2 s delay before the next trial for all stimulus conditions. After a trial finished (i.e. after either reward delivery or the end of the 2 s delay), an inter-trial interval of 1.5 to 2.5 s (uniform distribution) occurred before the software began to wait for the next quiescent period. Behavioural sessions were terminated at experimenter discretion once the mouse stopped performing the task (typically 1 h).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mice were trained in stages (Extended Data Fig. 1b). First, they were trained to ~70% performance with only coherent trials; then auditory, visual, and neutral/conflict trials were progressively introduced based on experimenter discretion. Using this training protocol, ~80% of mice learnt the task, and those that did learn reached the final stage in < 30 sessions (Extended Data Fig. 1c).

Behavioural quantification

For all analyses of behavioural data, timeouts and repeats following incorrect choices were excluded. To remove extended periods of mouse inattention at the start and end of experimental sessions, we excluded trials before/after the first/last three consecutive choices without a timeout. The 6% contrast level was included in analyses of inactivation experiments (Fig. 2) as all mice contributing to these analyses were presented with 6% contrast levels, but not all behavioural and electrophysiology sessions included 6% contrast.

On 91.5% of trials (142853/156118), subjects responded to the stimulus onset by turning the wheel within the 500 ms open-loop period (Extended Data Fig. 1f). For data analysis purposes, we therefore calculated mouse choice and reaction time from any wheel movements after stimulus onset (Extended Data Fig. 1d-e), even though during the task, rewards would only be delivered after the open-loop period had ended. These choices were defined by the first time point at which the movement exceeded ~30° of wheel rotation (the exact number varied across sessions/mice, Extended Data Fig. 1d), the same threshold required for reward delivery during the closed-loop period. This matched the outcome calculated during the closed loop period on 95.0% of trials (148203/156118). The reaction time was defined as the last time prior to the choice threshold at which velocity crossed 0 after at least 50 ms at 0 or opposite to the choice direction, and then exceeded 20% of the choice threshold per second for at least 50 ms (Extended Data Fig. 1e). On 5.1% of trials (8380/164498), no such timepoint existed or movement was non-zero within 10 ms of stimulus onset; these trials were excluded. On 38% of trials (59498/156118), mice made sub-threshold movements prior to their calculated reaction time. To eliminate the possibility that these earlier movements were responsible for the neural decoding of choice (Fig. 3g) we repeated this analysis using only trials without any movement prior to the calculated reaction time (Extended Data Fig. 4d), which did not change the results.

When calculating performance for each stimulus type for a single visual contrast (Fig. 1e, Extended Data Fig. 1i,k), the value for each mouse was calculated within each session before taking the mean across sessions. We then took the mean across symmetric presentations of each stimulus condition (e.g. unisensory auditory left and right trials). In the case of reaction time (Fig. 1b, Extended Data Fig. 1j,l), we calculated the median for each session before taking the mean across sessions and symmetric presentations; we also subtracted the mean across all stimulus types for each mouse. For both performance and reaction time, differences between stimulus types were quantified with a paired t-test (n = 17 mice). Using this analysis, we established that reaction times were faster on unisensory auditory trials than unisensory visual trials (Fig. 1b). To confirm that the earlier movements on unisensory auditory trials were genuine choices rather than reflexive movements unrelated to the stimulus location, we predicted whether stimuli were presented on the right or left in unisensory auditory and unisensory visual trials from the wheel velocity at each timepoint after stimulus onset. Trial data was subsampled for each session (to equalize the number of stimuli appearing on the left and right) and split into test and training data (2-fold cross validation). Mean prediction accuracy was calculated by first taking the mean across sessions, then across mice. Consistent with our conclusions from calculated reaction times, auditory location could be decoded earlier than visual location (Extended Data Fig. 1g). This conclusively demonstrates that mice were able to identify the location of an auditory stimulus earlier than a visual stimulus.

Video motion energy analysis

Because neural activity across the brain is related to bodily motion[61,62], we asked if mice still respond to stimuli in the passive condition. We filmed the mouse at 30 frames per second (DMK 23U618, The Imaging Source). We quantified the motion energy on each trial by averaging the absolute temporal difference in the pixel intensity values, across all pixels in a region of interest including the face and paws, and across a time period 0 to 400 ms after stimulus onset, which typically included the mouse response during behaviour (Extended Data Fig. 1d-f). This analysis established that mice exhibit minimal movement in response to task stimuli during passive conditions (Extended Data Fig. 5h).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Optogenetic inactivation

For optogenetic inactivation experiments (Fig. 2) we inactivated several cortical areas through the skull using a blue laser[55–58], in transgenic mice expressing ChR2 in Parvalbumin-expressing inhibitory interneurons (Ai32 x PV-Cre). Unilateral inactivation was achieved using a pair of mirrors mounted on galvo motors (GVSM002-EC/M, Thor labs) to orient the laser (L462P1400MM, Thor labs) to different points on the skull. On every trial, custom code drove the galvo motors to target one of 52 different coordinates distributed across the cortex (Fig. 2a), along with 2 control targets outside of the brain (Extended Data Fig. 3c). A 3D-printed isolation cone prevented laser light from reaching the screens and influencing behaviour. Inactivation coordinates were defined stereotaxically from bregma and were calibrated on each session. Anterior-posterior (AP) positions were distributed across 0, ±1, ±2, ±3, and -4 mm. Medial-lateral (ML) positions were distributed across ±0.6, ±1.8, ±3.0, and ±4.2 mm. On 75% of randomly interleaved trials, the laser (40 Hz sine wave, 462 nm, 3 mW) illuminated a pseudorandom location from stimulus onset until the end of the response window 1.5 s later (both open and closed loop periods, irrespective of mouse reaction time). The laser was not used on trial repetitions due to incorrect choices or timeouts. Pseudorandom illumination meant that a single cortical site was inactivated on only 1.4% of trials per session. This discouraged adaptation effects but required combining data across sessions for analyses. The galvo-mirrors were repositioned on every trial, irrespective of whether the laser was used, so auditory noise from the galvos did not predict inactivation.

Psychometric modelling

The additive behavioural model was fit in MATLAB using the *fmincon* function to iteratively optimise parameters using the *interior-point* algorithm to find 6 fit parameters: v_R ,v_L a_R and a_L representing sensitivities to right and left visual and auditory stimuli, b representing bias, and the contrast gain parameter γ. For individual mice (Fig. 1c,f, Extended Data Fig. 2a-o) models were fit to data combined across sessions, and we calculated the fraction of rightward choices across all trials. When the model was fit to combined data from multiple mice (Fig. 1d,g, Fig. 2f-h, Extended Data Fig. 3d-e), trials were subsampled to equalize numbers across mice before fitting the model. This subsampling process was repeated 10 times, and plots reflect the mean model parameters, and fraction of rightward choices, across repeats. For visualization, if the log-odds were not defined for a given stimulus condition (because a mouse, or mice, made only rightward or leftward choices) the log odds were regularized by adding one trial in each direction. This was only necessary for the coherent stimulus condition at 10% contrast in Fig. 2h.

We compared our additive behavioural model to a full model in which the log odds was computed for each stimulus combination, with a total of 25 parameters. We evaluated the fit of each model by its $\log_2$-likelihood ratio relative to a bias-only model $\log(p(R \mid A,V)/p(L \mid A,V)) = b$ using 5-fold cross-validation. After normalizing by the number of trials, this yields a quantity in bits per trial: the number of bits two parties would save in communicating the mouse's choice, if the stimulus is known to both. Across 17 mice, the additive model was not significantly worse than the full model, either when trained on all trial types (Fig. 1h), or when trained only on unisensory and neutral trials but tested on all trials including multisensory combinations (Extended Data Fig. 2p).

When fitting the additive model to data where different regions of dorsal cortex were inactivated, three target locations were combined for each dataset. For VISp (Fig. 2g), these were (-4,1.2), (-4,3) and (-3,3), where coordinates indicate (AP, ML) distances from bregma in mm. For MOs they were (2,0.6), (2,1.8) and (3,0.6) (Fig. 2h); for areas proximal to auditory cortex (-4,4.2), (-2,4.2), (-2,4.2) (Extended Data Fig. 3d); for areas proximal to somatosensory cortex (0,1.2), (0,3), (0,4.2) (Extended Data Fig. 3e). When fitting these models, the contrast gain parameter was fixed at the value obtained when fitting to non-inactivation trials.

Quantifying effects of optogenetic inactivation on choice

To quantify the change in the fraction of rightward choices when a particular cortical location was inactivated, we used a shuffle test (Fig. 2b-e, Extended Data Fig. 3a-b). Data were initially combined across 5 mice and segregated by stimulus type (unisensory visual, unisensory auditory, multisensory coherent, or multisensory conflict). For each type, data were further segregated into non-inactivation trials (laser off) and inactivation trials (laser on) grouped by the targeted area of dorsal cortex. For trials where the stimulus was presented on the right, we reversed the laterality of the stimulus and inactivation location such that all stimuli were effectively presented on the left (visual stimulus in the case of conflict trials). Data were randomly subsampled to equalize the number of trials contributed by each mouse

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to non-inactivation and the inactivation trials at each targeted location. We then calculated the difference in the fraction of rightward choices for each targeted location compared with non-inactivation trials. This process was repeated 25,000 times with different subsampling to produce a mean change in fraction of rightward choices for each inactivated location on dorsal cortex.

For each of the 25,000 iterations, we proceeded to generate 10 independent shuffles, where the labels for targeted location and trial identity (inactivation or non-inactivation) where randomly reassigned. We thus generated a null distribution for each targeted location, comprising 250,000 datapoints from independent shuffles. For each targeted location, the position of the unshuffled result within this null distribution gave the significance value for that location (e.g. top/bottom 0.05% for $p < 0.001$, top/bottom 0.005% for $p < 0.0001$).

When assessing the symmetry of inactivation effects across hemispheres (Extended Data Fig. 3a) the process was as described above, but without reversing the laterality of any trials. To confirm results were similar across mice (Extended Data Fig. 3b), we repeated this process for individual mice. In this case, the number of shuffled iterations remained at 250,000 but no subsampling was required (because there was no need to equalize across mice).

Quantifying effects of optogenetic inactivation on model parameters

To quantify the changes in parameters of the additive model (Fig. 2i, Extended Data Fig. 3f) the analysis closely mirrored the steps described above, but trial types were not segregated by stimulus type. The additive model was reparametrized such that stimuli were defined as being ipsilateral or contralateral to the site of inactivation, effectively combining data across hemispheres:

$$log\left(\frac{P(I)}{P(C)}\right) = b + v_i V_i^{\gamma} - v_c V_c^{\gamma} + a_i A_i - a_c A_c$$

Here, V_c and V_i are contralateral and ipsilateral visual contrasts, and A_c and A_i are contralateral and ipsilateral auditory azimuths. v_i ,v_c a_i and a_c represent sensitivities to contralateral and ipsilateral visual and auditory stimuli, while b represents the bias, and γ the contrast gain parameter. The unshuffled dataset comprised 2,500 different subsamples, and in each iteration, we fit the additive behavioural model to the non-inactivation data and to the inactivation data for each targeted location. This gave the mean change in each model parameter at each location on dorsal cortex. We compared this value to a null distribution (generated as described above, total of 25,000 independent shuffles) to establish the significance of each change. Since we observed no change in the contrast gain parameter, γ (Extended Data Fig. 3f), in our final analysis we fixed this value according to the non-inactivation trials and only quantified changes in the remaining 5 parameters (Fig. 2i).

The effect of inactivation on reaction time

We used a linear mixed effects model (LME) to determine the effect of inactivating VISp or MOs on mouse reaction time in the presence of a contralateral visual stimulus (Fig. 2j-k). For each mouse, we computed the median reaction time over trials of all sessions, for each combination of stimulus and inactivation conditions. We fit the following LME model to this data using MATLAB's *fitlme* function:

$$Reaction\ time \sim Inactivation + VisualContrast + (1|MouseID)$$

Here, $Reaction\ time$ is the response variable, $Inactivation$ (binary) and $VisualContrast$ (categorical) were fixed effect terms, and $MouseID$ was a random effect on the intercept. We fit four separate LMEs, one for each combination of coherent and conflict trials, and VISp and MOs inactivation. In each case, we assessed the sign and significance of the $Inactivation$ term to assess the impact of inactivation on mouse reaction time (Fig. 2j-k).

Neuropixels recordings

Recordings were made using Neuropixels (Phase3A; Ref.[81]) electrode arrays, which have 384 selectable recording sites out of 960 sites on a 1 cm shank. Probes were mounted to a custom holder (3D-printed polylactic acid piece) affixed to a steel rod held by a micromanipulator (uMP-4, Sensapex Inc.). Probes had a soldered external reference connected to ground which was subsequently connected to an Ag/AgCl wire positioned on the skull. On the first day of recording mice were briefly anaesthetized with isoflurane while one or two craniotomies were made with a biopsy punch. After

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

at least three hours of recovery, mice were head-fixed in the usual position. The craniotomies, as well as the ground wire, were covered with a saline bath. One or two probes were advanced through the dura, then lowered to their final position at approximately 10 µm/s.

Electrophysiological data were recorded with Open Ephys[82]. Raw data within the action potential band (1-pole high-pass filtered over 300 Hz) was denoised by common mode rejection (that is, subtracting the median across all channels), and spike-sorted using Kilosort 2 (www.github.com/MouseLand/Kilosort2). Units were manually curated using Phy to remove noise and multi-unit activity[83]. Each cluster of events ('unit') detected by a particular template was inspected, and if the spikes assigned to the unit resembled noise (zero or near-zero amplitude; non-physiological waveform shape or pattern of activity across channels), the unit was discarded. Units containing low-amplitude spikes, spikes with inconsistent waveform shapes, and/or refractory period contamination were labelled as 'multi-unit activity' and not included for further analysis.

To localize probe tracks histologically, probes were repeatedly dipped into a centrifuge tube containing DiI before insertion (ThermoFisher Vybrant V22888 or V22885). When probes were inserted along the same trajectory for multiple sessions (Extended Data Fig. 4a), they were coated with DiI on the first day, and subsequent recordings were estimated to have the same trajectory within the brain (although depth was independently estimated, Extended Data Fig. 4b). After experiments were concluded, mice were perfused with 4% paraformaldehyde. The brain was extracted and fixed for 24 h at 4 °C in paraformaldehyde before being transferred to 30% sucrose in PBS at 4 °C. The brain was then mounted on a microtome in dry ice and sectioned at 80 µm slice thickness. Sections were washed in PBS, mounted on glass adhesion slides, and stained with DAPI (Vector Laboratories, H-1500). Images were taken at 4× magnification for each section using a Zeiss AxioScan, in two colours: blue for DAPI and red for DiI. Probe trajectories were reconstructed from slice images (Extended Data Fig. 4a) using publicly available custom code (http://github.com/petersaj/AP_histology[84]). For each penetration, the point along the probe where it entered the brain was manually estimated using changes in the LFP signal (Extended Data Fig. 4b). Recordings were made in both left (47 penetrations) and right (41 penetrations) hemispheres. The position of each recorded unit within the brain was estimated from its depth along the probe. For visualization, the recorded cells were mapped onto a flattened cortex using custom code (Fig. 3a). Given the small size the frontal pole (FRP), neurons in this region could not be confidently separated from MOs, and so were considered part of MOs for the purpose of this manuscript (14% of MOs cells; excluding these cells did not significantly impact results).

Passive stimulus presentation recordings

Mice were presented with task stimuli under passive conditions after each behavioural recording session. Stimuli were presented in open-loop (entirely uncoupled from wheel movement) and mice did not receive rewards. Unisensory auditory, unisensory visual, coherent, and conflict trials were presented to mice. However, on unisensory visual trials, the auditory amplitude was set to zero (rather than positioned at 0° as in the task) to ensure visual sensory responses could be isolated. Due to time constraints, only one coherent and conflicting stimulus combination were presented (80% visual contrast in both cases), and the trial interval was reduced (randomly selected from 0.5 to 1 s). Stimulus conditions were randomly interleaved, and each condition was repeated ∼50 times.

Estimating firing rate

Unless otherwise specified, firing rates were calculated on each trial by binning in 2 ms windows and smoothing with a half-Gaussian filter with standard deviation of 60 ms. PSTHs were calculated by averaging this rate across trials.

Decoding stimuli and choices from population activity

To decode stimuli and choices from neural activity (Fig. 3e-g, Extended Data Fig. 4c-d), we trained a linear SVM decoder on the firing rate vector time-averaged over a window 0-300 ms after stimulus onset (Fig. 3e-f), 0-130 ms before movement onset (Fig. 3g, Extended Data Fig. 4d), or 150-300 ms after movement onset (Extended Data Fig. 4c). SVMs were trained separately for each Neuropixels behavioural recording, for any brain region with a minimum of 30 neurons recorded in that session. If more than 30 neurons were recorded, we repeatedly (5 repeats) selected a 30-neuron subset for decoding analysis and took the mean accuracy (5-fold cross-validated) across these repeats. Sessions with fewer than 25 trials of each decoded condition (e.g. left and right stimulus locations) were excluded. In the case of decoding visual location (Fig. 3e), only trials with high-contrast (40% and 80%) stimuli were included. In each session,

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

decoding accuracy was quantified as the fraction of test-set trials classified correctly, relative to the same number for a model with no access to the spike trains (whose optimal behaviour is to always predict the most common stimulus on the training set):

$$accuracy = \frac{neural\ decoding\ accuracy - baseline\ accuracy}{1 - baseline\ accuracy}$$

Combined-conditions choice/stimulus probability analysis

To quantify the selectivity of a cell for a choice while controlling for effects of stimulus (Extended Data Fig. 4f), we used the combined-conditions choice probability (ccCP) test introduced in Steinmetz *et al*[74]. This is based on an extension of the Mann-Whitney U statistic, defined as the fraction of pairs of trials of identical stimulus conditions but different choices, for which the firing rate on the right choice trial exceeds the firing rate on the left choice trial. The significance of this test statistic was evaluated by shuffling using a p-value of 0.01, meaning that the observed value has to be either below the 0.5 percentile or above the 99.5 percentile of a null distribution generated from 1000 shuffles of the choice labels for each stimulus condition in order to be deemed significant. For ccCP, we compared the firing rate averaged over 0 – 130 ms before movement onset between trials where the mouse made a leftward or rightward choice trials (Extended Data Fig. 4f).

To test for selectivity to one stimulus while controlling for the other stimulus and choice (Extended Data Fig. 4e), we used an analogous method, referred to as the combined conditions stimulus probability (ccSP). For visual ccSP, we compared the firing rate time-averaged over a 0 – 300 ms window after stimulus onset, between trials where the visual stimulus was on the left and trials where the visual stimulus was on the right, including only trials with high (40% or 80%) contrast (Extended Data Fig. 4e, left). For auditory ccSP, we compared the firing rate averaged over a time window 0 – 300 ms after stimulus onset between auditory-left and auditory-right trials (Extended Data Fig. 4e, right).

Modelling neural activity

To predict firing rate time courses from task events (Fig. 4a-c), we used an ANOVA-style decomposition. For this analysis, we pooled multisensory coherent and conflict trials of visual contrast 40% and 80% (using a single visual contrast did not impact results), resulting in four possible stimulus conditions: one for each combination of auditory and visual location. We defined binary variables $a_i, v_i, c_i = \pm 1$ encoding whether auditory stimuli, visual stimuli, and choices are to the left or right on trial i. We can decompose $\boldsymbol{F}_i(t)$, the firing rate vector on trial i at time t after stimulus onset, as:

$$\boldsymbol{F}_i(t) = \boldsymbol{B}(t) + a_i\boldsymbol{A}(t) + v_i\boldsymbol{V}(t) + a_iv_i\boldsymbol{N}(t) + \boldsymbol{M}(t-\tau_i) + c_i\boldsymbol{D}(t-\tau_i)$$

This model decomposes the response into a sum of 6 temporal kernels. $\boldsymbol{B}$ represents the grand mean stimulus response; $\boldsymbol{A}$ and $\boldsymbol{V}$ represent the additive main effects of auditory and visual stimulus location, and $\boldsymbol{N}$ represents a non-additive interaction between them. To account for the effects of movement, $\boldsymbol{M}$ is a kernel representing the mean effect of movement (relative to τ_i, the time of movement onset on trial i) and $\boldsymbol{D}$ represents the effect of movement direction. $\boldsymbol{B}, \boldsymbol{A}, \boldsymbol{V}, \boldsymbol{N}$ were allowed to be non-zero for -50 $\leq t \leq$ 400 ms. $\boldsymbol{M}, \boldsymbol{D}$ can be non-zero for -200 $\leq t - \tau_i \leq$ 700 ms. Only trials with $\tau_i < 300$ ms were included. The model was fit using ridge regression with a regularisation strength of $\alpha = 10$, which we found to give optimal prediction accuracy. We fit this model to each neuron in MOs with a non-zero firing rate during behaviour (n = 2183 neurons), using a training set consisting of half the trials (randomly selected). The error, E, of this fit was measured as:

$$E = \frac{1}{N_i}\frac{1}{N_t}\sum_i\sum_t(y_{it} - \hat{y}_{it})^2$$

Here, $\hat{y}_{it}$ and y_{it} are model prediction and test-set recorded firing rate on trial i and timepoint t, N_i is the number of neurons, and N_t is the number of time bins, spanning 0 to 400 ms relative to stimulus onset. E is thus the cross-validated mean-squared error between the predicted and the actual smoothed firing rate over this time window. To test for an additive code, we then repeated this process for an additive neural model where $\boldsymbol{N} = 0$ (Fig. 4c).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To investigate whether there was an interaction between stimulus condition and choice-related response, we also fit a model with 8 movement-aligned kernels, i.e. a movement and a direction kernel for each combination of the four possible audiovisual stimuli:

$$\boldsymbol{F}_i(t) = \boldsymbol{B}(t) + a_i\boldsymbol{A}(t) + v_i\boldsymbol{V}(t) + a_iv_i\boldsymbol{N}(t) + \boldsymbol{M}_{a_i,v_i}(t-\tau_i) + c_i\boldsymbol{D}_{a_i,v_i}(t-\tau_i)$$

We compared this full model to the additive neural model (two movement kernels and $\boldsymbol{N} = 0$) using the method described above (Extended Data Fig. 5d).

To model neural activity during passive stimulus presentation (Fig. 4d-g), we used a reduced model without movement-aligned kernels:

$$\boldsymbol{F}_i(t) = \boldsymbol{B}(t) + a_i\boldsymbol{A}(t) + v_i\boldsymbol{V}(t) + a_iv_i\boldsymbol{N}(t)$$

Here, only multisensory coherent and conflict trials of a single (80%) visual contrast were included (due to time constraints, this was the only contrast presented on multisensory trials in passive conditions). To test for an additive code, we repeated the process described above (on 2509 cells with non-zero firing rates (Fig. 4f). No regularization was used for this analysis of passive data as it did not improve fits.

To compare the fit of linear and non-linear models of neural firing (Fig. 4c,f, Extended Data Fig. 5d), we used a linear mixed effects method to determine the main effects of the prediction model, accounting for systematic differences in model fit across mice and across experiments within each mouse. This was done using the *fitlme* in MATLAB with the following formula:

$$error \sim 1 + model + (1 + model|subject) + (1 + model|subject{:}session) + (1|subject{:}session{:}neruon)$$

The error term E is modelled with an intercept, a fixed effect of the model type being used (e.g. either the additive or full model), random effects for the intercept and model type grouped by subjects, random effects for the intercept and model type grouped by session nested within subjects, and random effects for the intercept grouped by neurons nested within sessions within subjects. For all statistical tests we report the p-value of the main effect of the model type on the observed error values.

To examine the distribution of auditory and visual spatial sensitivity across neurons we used neural recordings from passive stimulus presentation (Fig. 4g, Extended Data Fig. 5e). We selected neurons where the additive neural model ($\boldsymbol{N} = 0$) explained a minimum of 2% variance. For each neuron, we averaged the amplitude of the $\boldsymbol{A}$ and $\boldsymbol{V}$ kernels over a time window from 0 to 300 ms after stimulus onset (the kernels were fit using all trials). To test for a significant correlation between the signed magnitude of these time-averaged $\boldsymbol{A}$ and $\boldsymbol{V}$ kernels, we used the linear mixed effects model described above, but with $time\text{-}averaged\ \boldsymbol{V}kernel$ and $time\text{-}averaged\ \boldsymbol{A}\ kernel$ substituted for $error$ and $model$ (Fig. 4g). To test for a relationship between the absolute values of the two kernels, we repeated this procedure but using the absolute, rather than the signed, time-averaged kernels (Extended Data Fig. 5e).

Lateralisation of stimulus and movement activity

To investigate whether there is lateralisation in the spatial preference of auditory neurons, we examined time-averaged value of the $\boldsymbol{A}$ kernels (0 ms to 300 ms after stimulus onset) after fitting the additive model ($\boldsymbol{N} = 0$) under passive conditions. We selected neurons for which the additive model performed better than a model with visual kernel alone, and compared the mean value of the $\boldsymbol{A}$ kernel for neurons recorded in each hemisphere (Extended Data Fig. 5b). We repeat the same procedure for the visual kernel weights to examine lateralisation of visual spatial preference (Extended Data Fig. 5a).

To investigate the lateralisation of movement-related responses, we repeated this procedure, but for the additive model ($\boldsymbol{N} = 0$) during behaviour. We then included all for which the directional movement kernel, $\boldsymbol{D}$ improved cross-validated fits. Mean kernel values of selected neurons were calculated using a time window -200 to 400 ms relative to movement onset (Extended Data Fig. 5c).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Statistical analysis to determine the lateralisation of sensory and movement responses were performed with linear mixed effects model as described above, but with $time\text{-}averaged\ kernel$ and $hemisphere$ substituted for $error$ and $model$.

Quantifying single-neuron discrimination time

To identify when visual and auditory information began to be encoded in MOs (Fig. 4h), we analysed responses to passive unisensory stimuli. We first used permutation tests to select neurons sensitive to the presence (On-Off) and/or the location (Right-Left) of auditory and visual stimuli. To identify On-Off neurons, we calculated two PSTHs, one for sounds in each location, in a window 0 to 300 ms after stimulus onset, and computed the difference between the maximum of this PSTH and the mean firing rate 300 to 0 ms before stimulus onset. We compared this value to a null distribution obtained from 1000 shufflings of the pre/post-stimulus windows independently for each trial. A neuron was defined as significantly responding to a stimulus if the maximum difference in unshuffled data was in the 1st or 99th percentile of the null distribution for either left or right stimuli. For Right-Left neurons, the same method was uses, but using the maximum difference between the PSTHs for left and right auditory or visual presentations 0 to 300 ms after stimulus onset, and shuffling the left/right trial labels. This method identified 72 auditory (3%) and 68 visual (3%) Right-Left neurons.

For identified On-Off neurons, we calculated the discrimination time by separately comparing the pre- and post-stimulus firing rate in a sliding window of 50 ms with step size 5 ms, defining significance using a Mann-Whitney U test at $p < 0.01$, and requiring three consecutive significant time windows to qualify as the discrimination time. We excluded discrimination times that occurred more than 300 ms after stimulus onset as they are unlikely to be stimulus-related activity. This analysis was done separately for left and right stimuli, taking the earliest statistically significant time window in either stimulus condition. For identified spatially selective (Right-Left) neurons, we defined the discrimination time as the earliest time after stimulus onset where there is a significant difference in the response to left and right stimuli (Fig. 4h). This method identified discrimination times for 82 and 36 auditory and visual On-Off neurons, and 59 and 36 auditory and visual Right-Left neurons. For each neuron we also calculated the 5-fold cross-validated decoding accuracy, relative to a baseline model (which always predicts the most-frequent stimulus-condition in the training set, as in Fig. 4e-f), from the time-averaged firing rate in a window 0 to 100ms after the discrimination time using a linear SVM decoder (Extended Data Fig. 5i).

Accumulator model

To investigate whether the structure of the sensory code in MOs can explain mouse behaviour, we fed this code into an accumulator model (Fig. 5, similar to a drift diffusion model[63]). Since stimulus responses were sparse in MOs (140 auditory or visual location-selective neurons total from all experiments, as defined by the criteria of the previous section, i.e. 6%), we combined neural activity across all mice and experiments. To do so, we first obtained the PSTH for each stimulus condition, from –100 ms to 300 ms relative to stimulus onset. We then simulated 360 "trials" per stimulus condition by generating surrogate spike trains from a Poisson process with intensity given by these PSTHs. The stimulus conditions include unisensory visual trials with contrasts of 10, 20, 40, and 80%, unisensory auditory trials, and coherent and conflict audiovisual trials where the visual contrast is at 80%. This process yielded a time-dependent rate vector $\boldsymbol{x}(t)$ for each trial, where t is time relative to stimulus onset.

The output of the accumulator model was a decision variable $d(t)$, produced by linearly accumulating neural activity:

$$d(t) = d(t-1) + \boldsymbol{x}(t) \cdot \boldsymbol{w}$$

Here, $\boldsymbol{w}$ is a set of time-independent weights that were learned by the model to optimize the speed and accuracy of its responses but were not fit to mouse behaviour. The choice of the model is defined by the sign of the decision variable when it crosses one of the thresholds: +1 or -1 for a rightward or leftward choice, and the reaction time of the model is the time of this threshold-crossing relative to stimulus onset (Fig. 5e).

To learn the weight vector $\boldsymbol{w}$, we define a target decision variable y for each trial, set to 1 or -1 for rightward or leftward stimuli on unisensory and coherent trials. On conflict trials, where there is no correct response, the target decision variable is randomly set to 1 or -1 with equal probability.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The weights were learned by minimising a loss function that compares the target decision variable with the model's output decision variable for each trial:

$$\mathcal{L}\big(d(t)\big) = \sum_{t<0} d(t)^2 + \sum_{t\geq 0} \max\big(0{,}1 - y\, d(t)\big)$$

We used a mean-squared error loss before the stimulus onset ($t < 0$), to ensure that the model does not make a decision before the stimulus onset. After the stimulus onset ($t \geq 0$) we use a hinge-loss error, which is zero when the decision variable is above the threshold for the correct choice, and penalizes incorrect decisions and decision values below the decision threshold. The loss function was minimised with respect to the weights of the model via gradient descent using the ADAM optimiser[85] with a learning rate of 0.01, and the gradient was obtained via automatic differentiation using the JAX library[86]. The model was trained on 70% of the trials for 300 epochs, and its behaviour was evaluated on the remaining 30% of the trials (Fig. 5c-d). To simulate the inactivation of the visual cortex in the right-hemisphere, we took the same learned model, but instead provided input where the activity of neurons that were previously identified as visual-left preferring neurons were decreased by 50% (Fig. 5d). During training, the decision boundaries were set to +1 and -1 respectively. To account for the choice bias that was observed in mice, we performed grid search on the decision boundary values after model training in order to minimise the mean-squared error between the choice probability observed in the mice and in the model averaged across all stimuli conditions (Fig. 5c-d).

Code & Data availability

The code used in the current study is available from the corresponding author on reasonable request. The datasets generated and/or analysed during the current study are available from the corresponding authors on reasonable request.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Appendix 1

Here we show that optimally combining information from two inputs with independent noise requires the log odds be an additive function of the two inputs.

Let $S \in \{L, R\}$ represent the (left/right) location of the stimulus. A prior estimate for this location is captured by a prior probability distribution $p(S)$. The two sensory inputs A and V follow conditional probability distributions $p(A \mid S)$ and $p(V \mid S)$. We assume them to be independent, so

$$p(V, A \mid S) = p(V \mid S)p(A \mid S).$$

By Bayes' theorem, the probability of the stimulus location given the inputs is:

$$p(S \mid V, A) = \frac{p(V, A \mid S)p(S)}{p(V, A)} = \frac{p(V \mid S)p(A \mid S)p(S)}{p(V, A)}$$

Let $p(R) = p(S = R)$ and $p(L) = p(S = L)$. Thus, the log odds is given by:

$$\log\left(\frac{p(R \mid V, A)}{p(L \mid V, A)}\right) = \log\left(\frac{p(V \mid R)p(A \mid R)p(R)/p(V, A)}{p(V \mid L)p(A \mid L)p(L)/p(V, A)}\right)$$

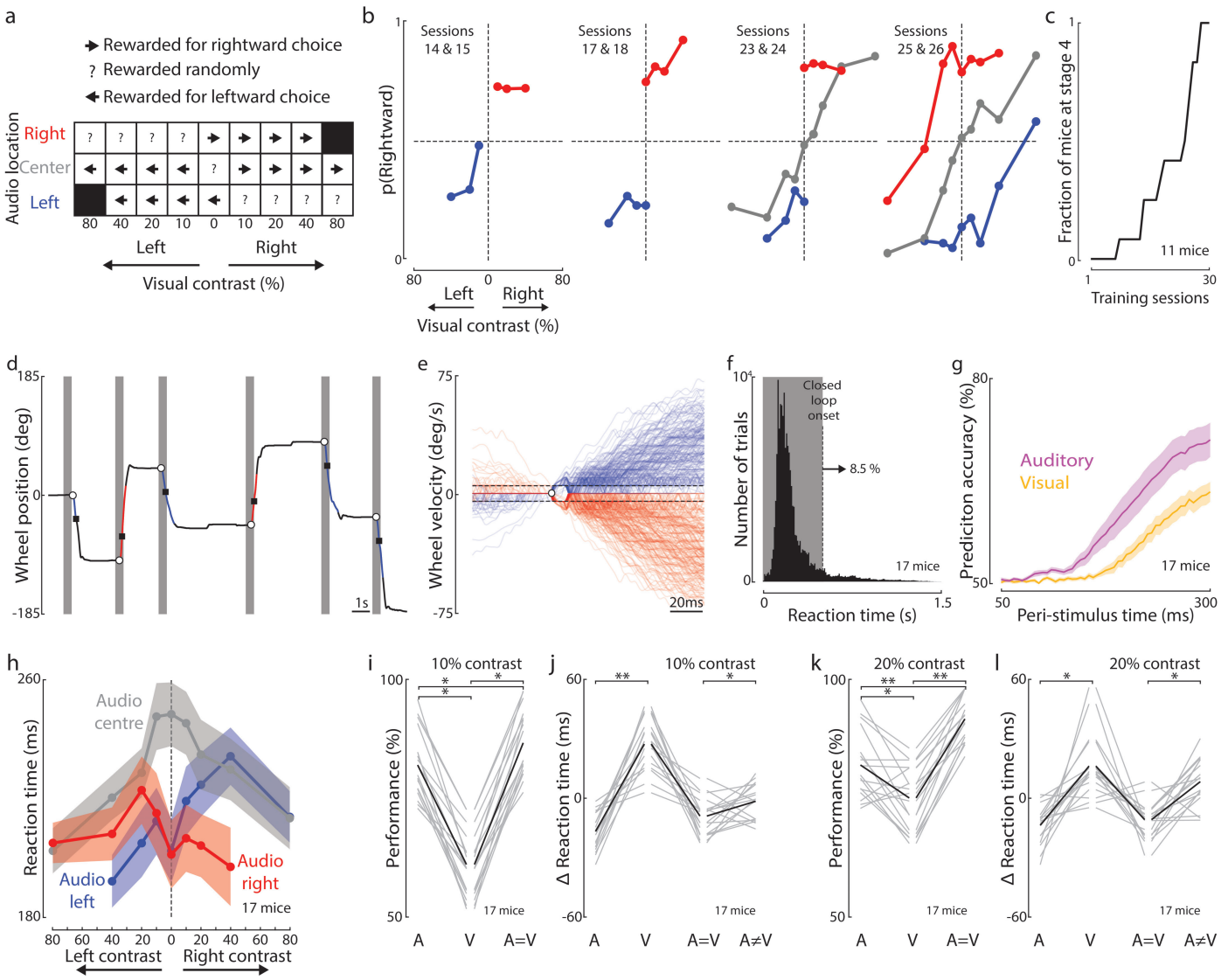
$$= \log\left(\frac{p(V \mid R)}{p(V \mid L)}\frac{p(A \mid R)}{p(A \mid L)}\frac{p(R)}{p(L)}\right)$$

$$= \log\left(\frac{p(V \mid R)}{p(V \mid L)}\right) + \log\left(\frac{p(A \mid R)}{p(A \mid L)}\right) + \log\left(\frac{p(R)}{p(L)}\right)$$

The first term depends only on V, the second only on A, while the third is a constant. Thus, the log odds is an additive function:

$$\log\left(\frac{p(R \mid V, A)}{p(L \mid V, A)}\right) = f(V) + g(A) + b$$

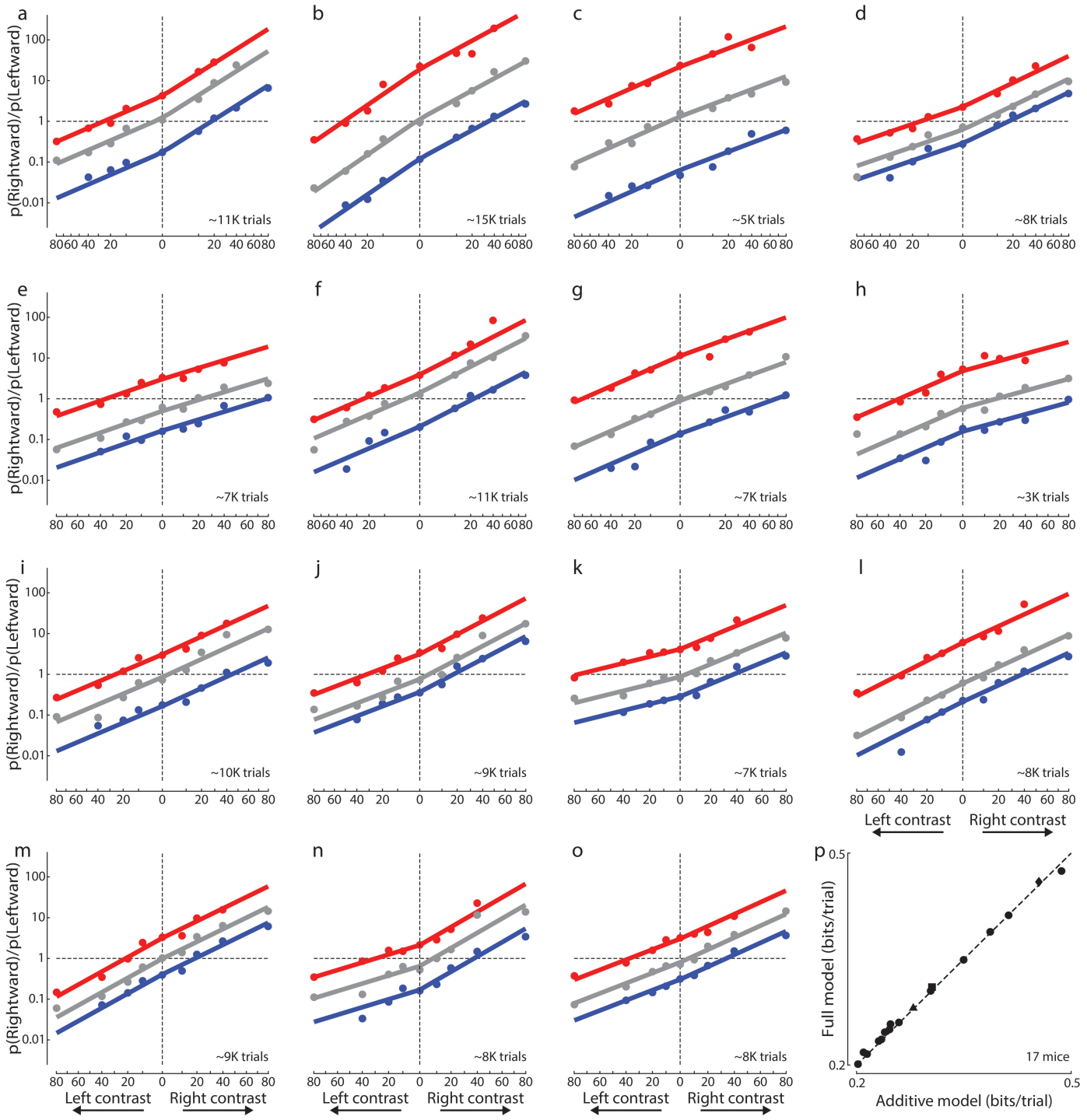
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Extended Data Fig. 1 Training and behavioural classification

(a) *Matrix of stimulus conditions and their corresponding reward criteria.* ***(b)*** *Training progression for one mouse. Mice first reach proficiency with audiovisual coherent trials (Stage 1), before we introduce unisensory auditory (Stage 2), unisensory visual (stage 3) and conflict (Stage 4) stimulus types. Plots for stages 1, 2, and 3, were made from the final two sessions of these stages. For stage 4, data is taken from the mouse's first two sessions at that stage.* ***(c)*** *Number of sessions required to train mice. In our final training pipeline, ~80% of mice learned the task. Those that did learn required less than 30 sessions.* ***(d)*** *Sample of wheel position trace from a single session. Grey regions indicate the 500ms open-loop period, beginning with stimulus onset. Open circles and squares indicate the movement onset and decision threshold (see Methods).* ***(e)*** *Zoom in to wheel velocity traces around the time of movement onset (open circle), for the behavioural session from (d). Red and blue lines indicate leftward and rightward choices. Dashed lines represent the velocity thresholds for movement onset (see Methods).* ***(f)*** *Histogram of reaction times across 17 mice (~156K trials). Shaded region indicates the 500ms open-loop period, during which 91.5% of movements were initiated.* ***(g)*** *Accuracy of predicting whether a stimulus was presented on the left or right side from wheel velocity at each time point relative to stimulus onset, using a threshold obtained by minimising an SVM loss function, for unisensory visual (gold) or unisensory auditory (magenta) trials. Shaded areas indicate the standard error across 17 mice. Earlier predictions on auditory trials confirm that mice can decode the location of auditory stimuli earlier than visual stimuli (i.e. earlier auditory reaction times do not reflect guesses).* ***(h)*** *The mean (across 17 mice) of the median reaction times at for each stimulus condition (~156K trials). Shading: standard error across mice.* ***(i)*** *Mouse performance (% rewarded trials) for unisensory auditory, unisensory visual, and coherent multisensory stimulus conditions (correct performance on conflict trials is undefined). This panel shows only trials with 10% visual contrast. Grey and black lines indicate individual mice and the mean across mice. $*$: $p < 10^{-6}$ (17 mice, paired t-test).* ***(j)*** *As in (i), but comparing median reaction times for each stimulus type relative to the mean across stimulus types for each mouse $**$: $p < 10^{-7}$, $*$: $p < 0.05$ (17 mice, paired t-test).* ***(k)*** *As in (i), but for trials with 20% visual contrast. $**$: $p < 10^{-7}$, $*$: $p < 0.01$ (17 mice, paired t-test).* ***(l)*** *As in (j), but for trials with 20% visual contrast. $*$: $p < 0.001$ (17 mice, paired t-test). Data with 40% contrast is shown in Fig. 1b,e.*

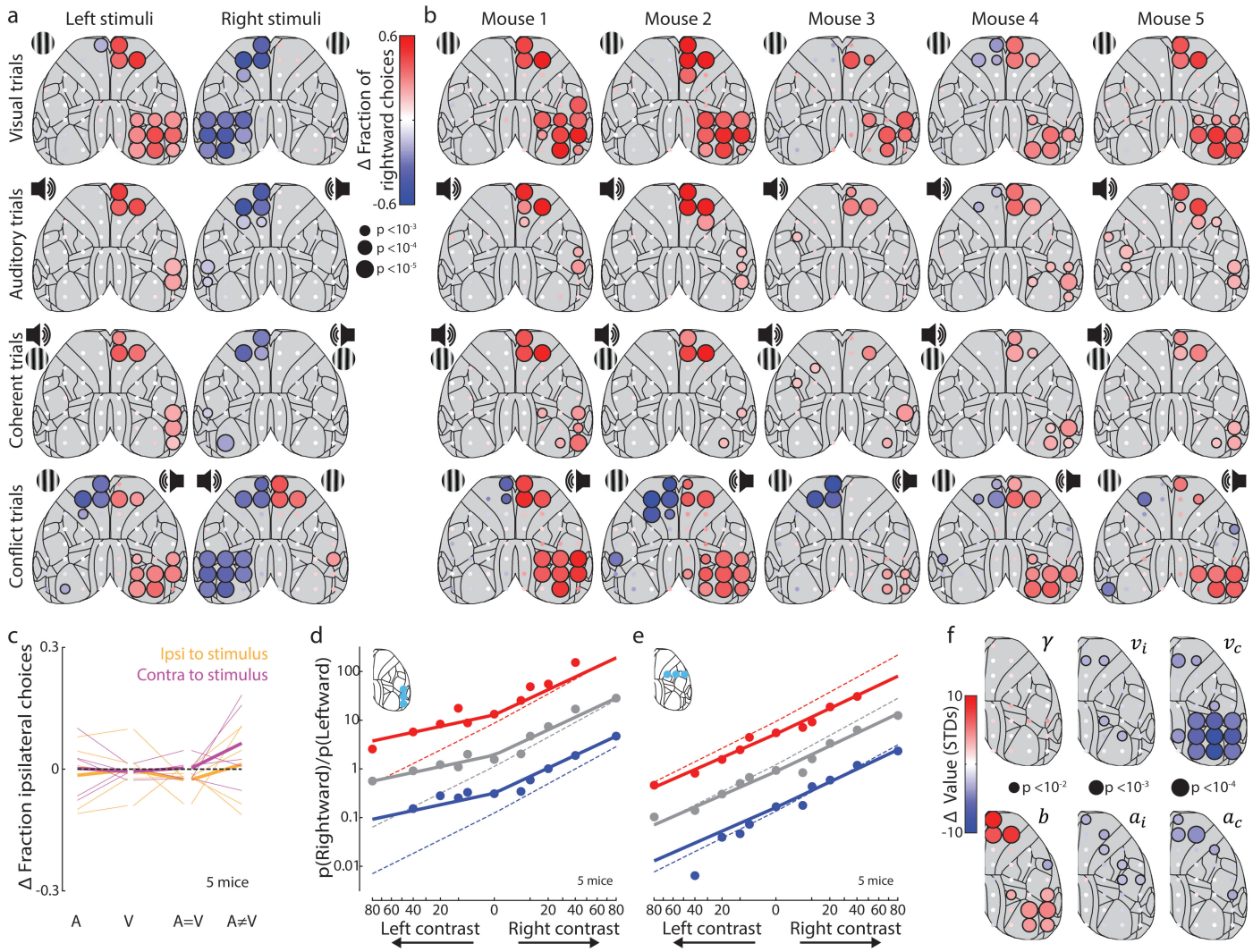
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Extended Data Fig. 2 Further evidence for additive model performance

(a-o) *Fits of the additive behavioural model to all 15 mice additional to those shown in Fig. 1c-d,f-g, with different auditory and visual proficiencies, plotted as in Fig. 1f.* ***(p)*** *Cross-validated log-likelihood ratio for the full model versus the additive behavioural model trained on only unisensory stimulus conditions, plotted as in Fig. 1h. There is no significant difference between models. $p > 0.05$ (17 mice, paired t-test).*

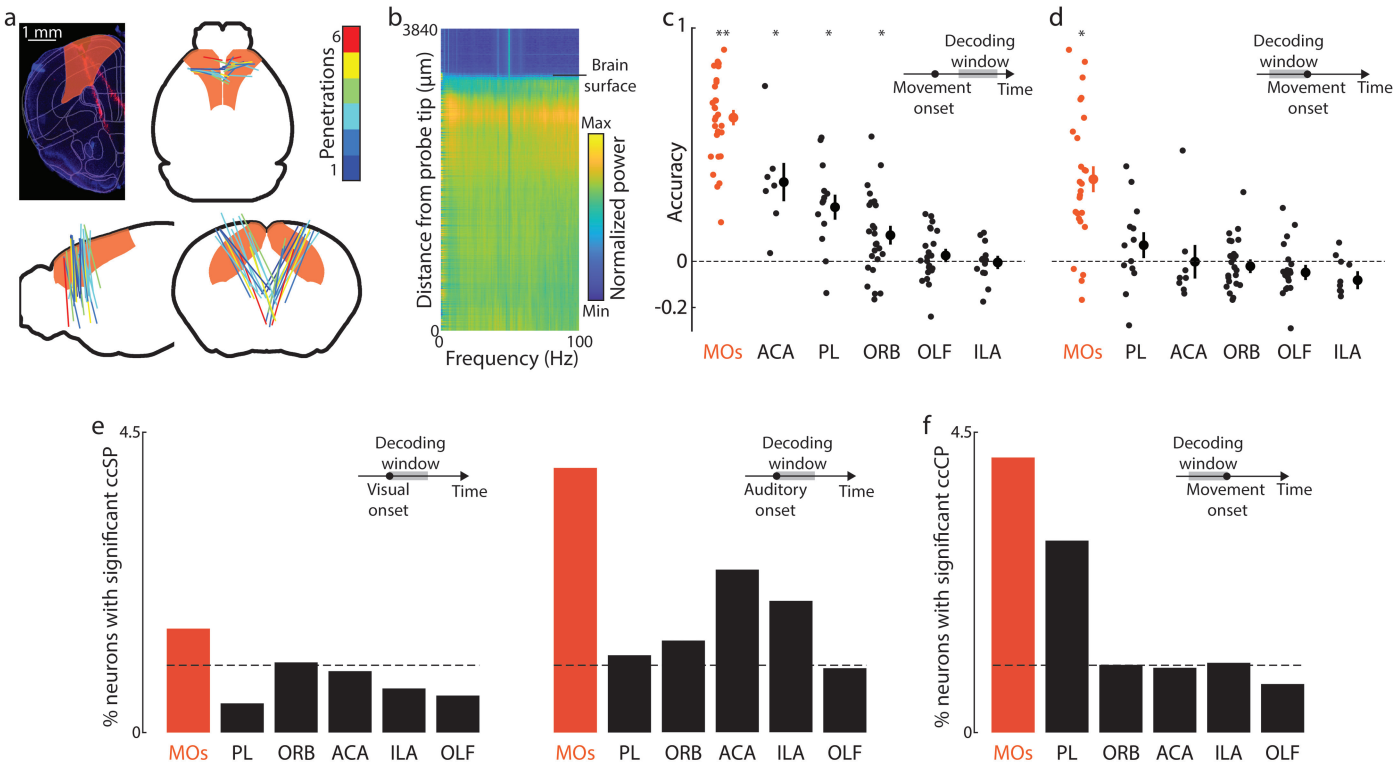
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Extended Data Fig. 3 Further data on optogenetic inactivation experiments

(a) *Change in the fraction of rightward choices for 52 stimulation sites for unisensory visual, unisensory auditory, coherent, and conflict stimulus conditions. The position of the grating and loudspeaker symbol indicate the locations of the visual and auditory stimuli in each case; Left and right columns represent mirrored cases of the same stimulus condition. Dot colour indicates change in the fraction of rightward choices and dot size represents statistical significance (5 mice, permutation test, see Methods).* ***(b)*** *As in (a), but each column represents an individual mouse, and trials were combined and reversed such that the stimulus (visual in conflict trials) was on the left. Dot colour indicates change in the fraction of rightward choices.* ***(c)*** *Change in the fraction of ipsilateral choices on trials when inactivation target was outside the brain. No significant effect was observed for any stimulus condition for ipsilateral (gold) or contralateral (magenta) inactivation sites. $p > 0.05$ (paired t-test).* ***(d)*** *Fit of the additive behavioural model to trials in which any of 3 sites proximal to right auditory cortex were inactivated. Red/grey/blue lines (fit) and points (data) indicate trials where the auditory stimulus was presented on the right/centre/left (5 mice, 6600 trials). Dashed lines indicate model fit to non-inactivation trials. The proximity of these regions to visual cortex likely explains their effects on contralateral visual stimuli. Trials with inactivation of 3 sites proximal to left auditory cortex were also included in the average after reflection.* ***(e)*** *As in (d), but for trials when somatosensory cortex was inactivated (5 mice, 6689 trials).* ***(f)*** *As in main text Fig. 2i, but also allowing the contrast gain saturation parameter y to be fit. No significant changes in the contrast gain parameter (γ) were observed (5 mice, permutation test, see Methods).*

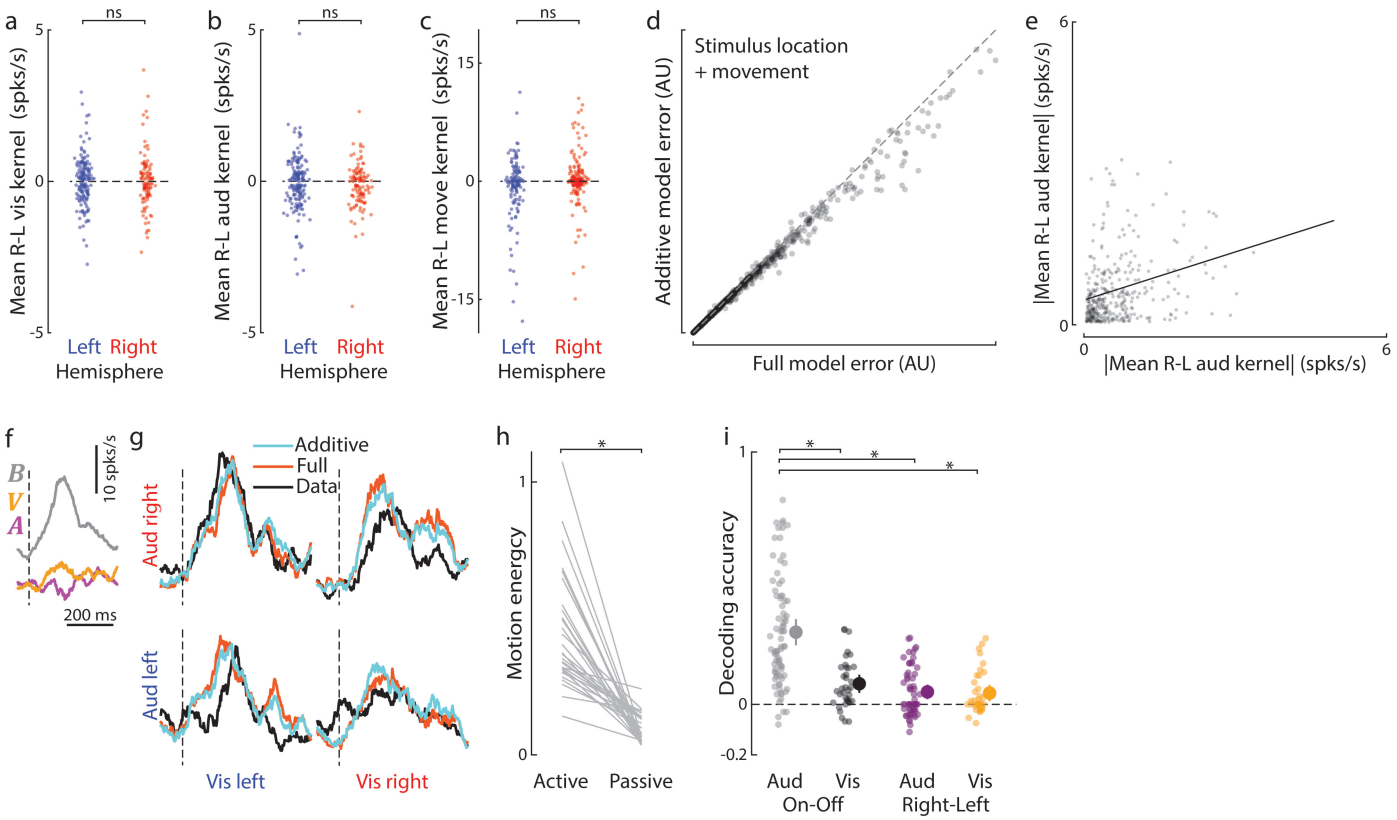
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Extended Data Fig. 4 Electrophysiology methods and controls

(a) *Histological reconstruction of electrode tracks. Top left: Example coronal section (DAPI-staining, blue) at 1.86mm anterior to bregma with DiI-labelled probe (red) tracks from 2 insertions. Overlay of Allen atlas brain regions[87], with MOs highlighted in orange. Slices were registered to Allen Atlas using custom software (see Methods). Right and bottom: reconstructed trajectories of all recordings analysed, projected onto coronal, sagittal, and horizontal planes. Colour indicates the number of penetrations along a given trajectory (one DiI-labelled penetration, same angles and point of insertion).* ***(b)*** *Normalized power spectrum (log(power), z-scored across probe depth) versus depth along probe for a single penetration. The point where the probe enters the brain can be identified by a sudden increase in power.* ***(c)*** *Cross-validated accuracy (relative to a bias model, see Methods) of an SVM decoder trained to predict rightward or leftward choices based on population spiking activity time-averaged over a window 150 ms to 300 ms after movement onset. Each point represents the decoding accuracy from neurons in one brain region (Secondary motor (MOs), orbitofrontal (ORB), anterior cingulate (ACA), prelimbic (PL), infralimbic (ILA), or Olfactory (OLF)), from a single experimental session. For each point, 30 neurons were randomly selected to equalize population size across sessions and brain regions. $**$: $p < 10^{-6}$, $*$: $p < 0.01$. ($\geq$ 7 sessions from 3-6 mice for each region, t-test).* ***(d)*** *As in (c), but with a time window of 0 ms to 130 ms preceding movement. Trials where mice made sub-threshold movements prior to movement onset were excluded (see Methods). $**$: $p < 10^{-6}$.* ***(e)*** *The proportion of neurons sensitive to visual (left) or auditory (right) stimulus location, estimated with combined conditions stimulus probability analysis (see Methods), after controlling for the other stimulus and for choice, using neural activity time-averaged over a window 0 to 300 ms after stimulus onset. In both cases, MOs has the highest proportion of significant neurons ($\geq$ 700 neurons per area).* ***(f)*** *As in (e), but using combined conditions choice probability (see Methods) to estimate the proportion of neurons sensitive to the upcoming choice, after controlling for both stimuli, using neural activity time-averaged over a window 0 to 130 ms before movement onset.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Extended Data Fig. 5 Additional analysis for additive data

(a) *Time average of the visual R-L kernel ($\boldsymbol{V}, Eqn\ 5$, 0 ms to 300 ms after stimulus onset) versus recorded hemisphere, after fitting the additive neural model ($\boldsymbol{N} = 0$) under passive conditions ($\boldsymbol{D}, \boldsymbol{M} = 0$) (see Methods). There is no significant lateralization in spatial preference ($p > 0.05$; 273 cells, linear mixed-effects model).* ***(b)*** *As in (a), but for the auditory R-L kernel ($\boldsymbol{A}$). ($p > 0.05$; 287 cells linear mixed-effects model).* ***(c)*** *As in (a), but the movement R-L kernel ($\boldsymbol{D}$, -200 to 400 ms relative to movement onset), after fitting the additive model during active behaviour ($p > 0.05$; 209 cells, linear mixed-effects model).* ***(d)*** *Prediction error for each neuron for the additive and full sensory-movement models (see Methods). The full model has significantly higher error. ($p < 0.01$, 2183 cells, linear mixed-effects model). For visualization purposes, the largest 1% of errors were excluded from the plot, but not from statistical analysis.* ***(e)*** *Absolute value of time-averaged visual R-L kernel ($\boldsymbol{V}$), versus absolute value of time-averaged auditory R-L kernel ($\boldsymbol{A}$) versus recorded hemisphere, after fitting the additive model under passive conditions for significant neurons. A time window from 0 to 300 ms after stimulus onset was used in both cases. Absolute responses are correlated, suggesting that if a neuron is predictable from spatial stimuli in one modality, it is more likely to be predictable from the other modality ($p < 10^{-5}$, 2509 cells, linear mixed-effects model).* ***(f)*** *Example sensory kernels from fitting the additive neural model to a single neuron under passive conditions. The selected neuron has opposing sensitivities for auditory (left-preference) and visual (right-preference) stimulus locations.* ***(g)*** *Cross-validated fit of the model kernels from (f) to neural responses under passive conditions for all audiovisual combinations. Cyan and orange lines show predictions of the additive and full models, black line shows test-set average responses.* ***(h)*** *Difference in facial video motion energy (see Methods) when mice are performing the behaviour (active) versus passive presentation of stimuli (passive). Mice exhibit significantly less motion under passive conditions. $*$: $p < 10^{-8}$ (30 sessions, paired t-test).* ***(i)*** *Neurons were tested individually to see if they encoded the presence or location of visual and auditory stimuli (see Methods). The plot shows the single-neuron decoding accuracy of all significant neurons for each discrimination. Neurons encoding auditory stimulus presence (grey, 82 cells) have higher decoding accuracy than all other categories (36/59/36 cells for black/magenta/gold). $*$: $p < 10^{-6}$ (t-test).*

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References:

1. Gold, J. I. & Shadlen, M. N. Banburismus and the Brain Decoding the Relationship between Sensory Stimuli, Decisions, and Reward. *Neuron* **36**, 299–308 (2002).

2. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429 (2002).

3. Angelaki, D. E., Gu, Y. & DeAngelis, G. C. Multisensory integration: psychophysics, neurophysiology, and computation. *Curr Opin Neurobiol* **19**, 452–458 (2009).

4. Chandrasekaran, C. Computational principles and models of multisensory integration. *Curr Opin Neurobiol* **43**, 25–34 (2017).

5. Gu, Y., Angelaki, D. E. & DeAngelis, G. C. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci* **11**, 1201–1210 (2008).

6. Fetsch, C. R., Pouget, A., DeAngelis, G. C. & Angelaki, D. E. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci* **15**, 146–154 (2012).

7. Alais, D. & Burr, D. The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology* **14**, 257–62 (2004).

8. Raposo, D., Sheppard, J. P., Schrater, P. R. & Churchland, A. K. Multisensory Decision-Making in Rats and Humans. *J Neurosci* **32**, 3726–3735 (2012).

9. Witten, I. B. & Knudsen, E. I. Why Seeing Is Believing: Merging Auditory and Visual Worlds. *Neuron* **48**, 489–496 (2005).

10. Colavita, F. B. Human sensory dominance. *Percept Psychophys* **16**, 409–412 (1974).

11. Sinnett, S., Spence, C. & Soto-Faraco, S. Visual dominance and attention: The Colavita effect revisited. *Percept Psychophys* **69**, 673–686 (2007).

12. Battaglia, P. W., Jacobs, R. A. & Aslin, R. N. Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am* **20**, 1391 (2003).

13. Fetsch, C. R., Turner, A. H., DeAngelis, G. C. & Angelaki, D. E. Dynamic Reweighting of Visual and Vestibular Cues during Self-Motion Perception. *J Neurosci* **29**, 15601–15612 (2009).

14. Song, Y.-H. *et al.* A Neural Circuit for Auditory Dominance over Visual Perception. *Neuron* **93**, 940-954.e6 (2017).

15. Gharaei, S., Arabzadeh, E. & Solomon, S. G. Integration of visual and whisker signals in rat superior colliculus. *Sci Rep-uk* **8**, 16445 (2018).

16. Costa, M., Piché, M., Lepore, F. & Guillemot, J.-P. Age-related audiovisual interactions in the superior colliculus of the rat. *Neuroscience* **320**, 19–29 (2016).

17. King, A. & Palmer, A. Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research* **60**, (1985).

18. Stein, B. E. & Stanford, T. R. Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience* **9**, 255–266 (2008).

19. Meredith, M. & Stein, B. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of neurophysiology* **56**, 640–62 (1986).

20. Lohse, M., Dahmen, J. C., Bajo, V. M. & King, A. J. Subcortical Circuits Mediate Communication Between Primary Sensory Cortical Areas. *Biorxiv* 2020.04.28.064857 (2020) doi:10.1101/2020.04.28.064857.

21. Bieler, M., Xu, X., Marquardt, A. & Hanganu-Opatz, I. L. Multisensory integration in rodent tactile but not visual thalamus. *Sci Rep-uk* **8**, 15684 (2018).

22. Chou, X. *et al.* Contextual and cross-modality modulation of auditory cortical processing through pulvinar mediated suppression. *Elife* **9**, e54157 (2020).

23. Komura, Y., Tamura, R., Uwano, T., Nishijo, H. & Ono, T. Auditory thalamus integrates visual inputs into behavioral gains. *Nat Neurosci* **8**, 1203–1209 (2005).

24. Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. *Nature Neuroscience* **17**, (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

25. Lippert, M. T., Takagaki, K., Kayser, C. & Ohl, F. W. Asymmetric multisensory interactions of visual and somatosensory responses in a region of the rat parietal cortex. *PloS one* **8**, e63631 (2013).

26. Nikbakht, N., Tafreshiha, A., Zoccolan, D. & Diamond, M. E. Supralinear and Supramodal Integration of Visual and Tactile Signals in Rats: Psychophysics and Neuronal Mechanisms. *Neuron* **97**, 626-639.e8 (2018).

27. Ohshiro, T., Angelaki, D. E. & DeAngelis, G. C. A normalization model of multisensory integration. *Nature neuroscience* **14**, 775–82 (2011).

28. Rohe, T. & Noppeney, U. Distinct Computational Principles Govern Multisensory Integration in Primary Sensory and Association Cortices. *Curr Biol* **26**, 509–514 (2016).

29. Khandhadia, A. P., Murphy, A. P., Romanski, L. M., Bizley, J. K. & Leopold, D. A. Audiovisual integration in macaque face patch neurons. *Curr Biol* (2021) doi:10.1016/j.cub.2021.01.102.

30. Avillac, M., Hamed, S. B. & Duhamel, J.-R. Multisensory Integration in the Ventral Intraparietal Area of the Macaque Monkey. *J Neurosci* **27**, 1922–1932 (2007).

31. Seilheimer, R. L., Rosenberg, A. & Angelaki, D. E. Models and processes of multisensory cue combination. *Current Opinion in Neurobiology* **25**, 3846 (2013).

32. Avillac, M., Denève, S., Olivier, E., Pouget, A. & Duhamel, J.-R. Reference frames for representing visual and tactile locations in parietal cortex. *Nat Neurosci* **8**, 941–949 (2005).

33. Cao, Y., Summerfield, C., Park, H., Giordano, B. L. & Kayser, C. Causal Inference in the Multisensory Brain. *Neuron* **102**, 1076-1087.e8 (2019).

34. Gu, Y., Cheng, Z., Yang, L., DeAngelis, G. C. & Angelaki, D. E. Multisensory Convergence of Visual and Vestibular Heading Cues in the Pursuit Area of the Frontal Eye Field. *Cereb Cortex* **26**, 3785–3801 (2015).

35. Jones, E. G. & Powell, T. P. S. An Anatomical Study of Converging Sensory Pathways Within the Cerebral Cortex of the Monkey. *Brain* **93**, 793–820 (1970).

36. Iurilli, G. *et al.* Sound-Driven Synaptic Inhibition in Primary Visual Cortex. *Neuron* **73**, 814–828 (2012).

37. Meijer, G. T., Montijn, J. S., Pennartz, C. M. & Lansink, C. S. Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency. *The Journal of Neuroscience* **37**, 8783–8796 (2017).

38. Ibrahim, L. A. *et al.* Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition. *Neuron* **89**, 1031–1045 (2016).

39. Bizley, J. K. & King, A. J. Visual influences on ferret auditory cortex. *Hearing Research* **258**, (2009).

40. Bizley, J. K. & King, A. J. Visual–auditory spatial processing in auditory cortical neurons. *Brain Research* **1242**, 24–36 (2008).

41. Atilgan, H. *et al.* Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron* **97**, 640-655.e4 (2018).

42. Ghazanfar, A. A. & Schroeder, C. E. Is neocortex essentially multisensory? *Trends Cogn Sci* **10**, 278–285 (2006).

43. Driver, J. & Noesselt, T. Multisensory Interplay Reveals Crossmodal Influences on 'Sensory-Specific' Brain Regions, Neural Responses, and Judgments. *Neuron* **57**, 11–23 (2008).

44. Meredith, M. A. & Allman, B. L. Single-unit analysis of somatosensory processing in the core auditory cortex of hearing ferrets. *Eur J Neurosci* **41**, 686–698 (2015).

45. Kayser, C., Petkov, C. I. & Logothetis, N. K. Visual Modulation of Neurons in Auditory Cortex. *Cerebral Cortex* **18**, 1560–1574 (2008).

46. Licata, A. M. *et al.* Posterior Parietal Cortex Guides Visual Decisions in Rats. *J Neurosci* **37**, 4954–4966 (2017).

47. Gu, Y., DeAngelis, G. C. & Angelaki, D. E. Causal Links between Dorsal Medial Superior Temporal Area Neurons and Multisensory Heading Perception. *J Neurosci* **32**, 2299–2313 (2012).

48. Bizley, J. K., Jones, G. P. & Town, S. M. Where are multisensory signals combined for perceptual decision-making? *Curr Opin Neurobiol* **40**, 31–37 (2016).

49. Burgess, C. P. *et al.* High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Reports* **20**, 2513–2524 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

50. Meijer, G. T., Pie, J. L., Dolman, T. L., Pennartz, C. M. A. & Lansink, C. S. Audiovisual Integration Enhances Stimulus Detection Performance in Mice. *Front Behav Neurosci* **12**, 231 (2018).

51. Hammond-Kenny, A., Bajo, V. M., King, A. J. & Nodal, F. R. Behavioural benefits of multisensory processing in ferrets. *Eur J Neurosci* **45**, 278–289 (2017).

52. Raab, D. H. Division of psychology: Statistical facilitation of simple reaction times. *T New York Acad Sci* **24**, 574–590 (1962).

53. Stein, B. E., Meredith, A. M., Huneycutt, S. W. & McDade, L. Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. *Journal of Cognitive Neuroscience* **1**, 12–24 (1989).

54. Pisupati, S., Chartarifsky-Lynn, L., Khanal, A. & Churchland, A. K. Lapses in perceptual decisions reflect exploration. *Elife* **10**, e55490 (2021).

55. Zatka-Haas, P., Steinmetz, N. A., Carandini, M. & Harris, K. D. Distinct contributions of mouse cortical areas to visual discrimination. *bioRxiv* 501627 (2019) doi:10.1101/501627.

56. Guo, Z. V. *et al.* Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81**, 179–194 (2014).

57. Cardin, J. A. *et al.* Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459**, 663–667 (2009).

58. Olsen, S. R., Bortone, D. S., Adesnik, H. & Scanziani, M. Gain control by layer six in cortical circuits of vision. *Nature* **483**, 47–52 (2012).

59. Glickfeld, L. L., Histed, M. H. & Maunsell, J. H. R. Mouse Primary Visual Cortex Is Used to Detect Both Orientation and Contrast Changes. *J Neurosci* **33**, 19416–19422 (2013).

60. Jenkins, W. M. & Merzenich, M. M. Role of cat primary auditory cortex for sound-localization behavior. *J Neurophysiol* **52**, 819–847 (1984).

61. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, eaav7893 (2019).

62. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci* **22**, 1677–1686 (2019).

63. Ratcliff, R. A theory of memory retrieval. *Psychol Rev* **85**, 59–108 (1978).

64. Gold, J. I. & Shadlen, M. N. The Neural Basis of Decision Making. *Annual Review of Neuroscience* **30**, 535–574 (2007).

65. Ebbesen, C. L. *et al.* More than Just a "Motor": Recent Surprises from the Frontal Cortex. *J Neurosci* **38**, 9402–9413 (2018).

66. Barthas, F. & Kwan, A. C. Secondary Motor Cortex: Where 'Sensory' Meets 'Motor' in the Rodent Frontal Cortex. *Trends Neurosci* **40**, 181–193 (2017).

67. Duan, C. A., Erlich, J. C. & Brody, C. D. Requirement of Prefrontal and Midbrain Regions for Rapid Executive Control of Behavior in the Rat. *Neuron* **86**, 1491–1503 (2015).

68. Hanks, T. D. *et al.* Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* **520**, 220 (2015).

69. Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D. & Brody, C. D. Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *Elife* **4**, e05457 (2015).

70. Erlich, J. C., Bialek, M. & Brody, C. D. A Cortical Substrate for Memory-Guided Orienting in the Rat. *Neuron* **72**, 330–343 (2011).

71. Insanally, M. N. *et al.* Spike-timing-dependent ensemble encoding by non-classically responsive cortical neurons. *Elife* **8**, e42409 (2019).

72. Sul, J. H., Jo, S., Lee, D. & Jung, M. W. Role of rodent secondary motor cortex in value-based action selection. *Nat Neurosci* **14**, 1202–1208 (2011).

73. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat Neurosci* **9**, 1432–1438 (2006).

74. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.26.441250; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

75. Maass, W., Natschläger, T. & Markram, H. Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. *Neural Comput* **14**, 2531–2560 (2002).

76. Caron, S. J. C., Ruta, V., Abbott, L. F. & Axel, R. Random convergence of olfactory inputs in the Drosophila mushroom body. *Nature* **497**, 113–117 (2013).

77. Li, N., Chen, T.-W., Guo, Z. V., Gerfen, C. R. & Svoboda, K. A motor cortex circuit for motor planning and movement. *Nature* **519**, 51–56 (2015).

78. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).

79. Steinmetz, N. A. *et al.* Aberrant Cortical Activity in Multiple GCaMP6-Expressing Transgenic Mouse Lines. *Eneuro* **4**, ENEURO.0207-17.2017 (2017).

80. Bhagat, J., Wells, M. J., Harris, K. D., Carandini, M. & Burgess, C. P. Rigbox: An Open-Source Toolbox for Probing Neurons and Behavior. *Eneuro* **7**, ENEURO.0406-19.2020 (2020).

81. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232 (2017).

82. Siegle, J. H. *et al.* Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology. *J Neural Eng* **14**, 045003 (2017).

83. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat Neurosci* **19**, 634–641 (2016).

84. Peters, A. J., Fabre, J. M. J., Steinmetz, N. A., Harris, K. D. & Carandini, M. Striatal activity topographically reflects cortical activity. *Nature* **591**, 420–425 (2021).

85. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *Arxiv* (2014).

86. Frostig, R., Johnson, M. J. & Leary, C. Compiling machine learning programs via high-level tracing. *1st SysML Conference* (2018).

87. Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**, 936-953.e20 (2020).