Cortical circuits for cross-modal generalization

Maëlle Guyoton[1]*, Giulio Matteucci[1]*, Charlie G. Foucher[1], Matthew P. Getz[2], Julijana Gjorgjieva[2] & Sami El-Boustani[1]

[1] Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, 1 Rue Michel-Servet, 1206 Geneva, Switzerland.

[2] School of Life Sciences, Technical University of Munich, 85354 Freising, Germany.

* These authors contributed equally to this work.

Summary

Adapting goal-directed behaviors to changing sensory conditions is a fundamental aspect of intelligence. The brain uses abstract representations of the environment to generalize learned associations across sensory modalities. The circuit organization that mediates such cross-modal generalizations remains, however, unknown. Here, we demonstrate that mice can bidirectionally generalize sensorimotor task rules between touch and vision by using abstract representations of peri-personal space within the cortex. Using large-scale mapping in the dorsal cortex at single-cell resolution, we discovered multimodal neurons with congruent spatial representations within multiple associative areas of the dorsal and ventral stream. Optogenetic sensory substitution and systematic silencing of these associative areas revealed that a single area in the dorsal stream is necessary and sufficient for cross-modal generalization. Our results identify and comprehensively describe a cortical circuit organization that underlies an essential cognitive function, providing a structural and functional basis for abstract reasoning in the mammalian brain.

Introduction

Objects possess unique physical properties that are detected by different sensory organs. The brain seamlessly integrates these sensory inputs to create unified percepts and abstract representations of the environment, which are essential for generalizing behaviors to unfamiliar situations[1]. This phenomenon is especially pronounced in the peri-personal space, where visual and tactile systems converge[2]. Indeed, an object's position and identity, initially discerned through touch in darkness, can immediately be recognized by sight in light. Cross-modal generalization —also called cross-modal transfer learning— describes the process by which recognition in one sensory modality enables the generalization of learned associations to others, a capability observed across diverse species[2–5], suggesting a foundational circuit organization yet to be elucidated. While neuronal correlates of increasingly abstract representations have been reported[6–9], the cortical architecture enabling the use of such representations for generalized learning across sensory modalities is not yet understood. Mice have proven to be a valuable model for dissecting circuits responsible for multisensory integration and their role in goal-directed behavior[10–13]. Mice rely on visuo-tactile inputs for behaviors such as gap crossing[14], navigation[15], and object recognition[6]. Because whiskers occupy a substantial portion of the visual field, both somatosensory and visual systems frequently receive correlated inputs (Fig. 1a). In the superior colliculus, multimodal representations of whisker and visual information possess topographically co-aligned functional maps[16,17], potentially enhancing reflexive behaviors like gaze and head orientation by increasing the salience of spatially congruent multisensory events[18]. Cortical circuits, critical for conscious perception and goal-directed behaviors, also display multimodal visuo-tactile responses. In particular, associative areas within the posterior parietal cortex

receive inputs from both the primary visual cortex (V1) and the primary somatosensory cortex (S1)[19]. However, the precise nature of the functional and anatomical organization of cortical areas dedicated to visuo-tactile processing, as well as their potential role in cross-modal generalization of previously learned sensorimotor associations, remains unknown. Here we investigate these circuits in mice performing goal-directed behaviors.

Results

Cross-modal generalization in mice

We designed a behavioral paradigm to test the ability of mice to generalize sensorimotor associations learned through whisker sensations to the visual modality. Given that both sensory modalities share a common spatial organization within the peri-personal space (Fig. 1a), the task includes visual and tactile stimuli that originate from locations that are spatially congruent (Fig. 1b). In the dark, head-fixed and water-restricted mice were first trained on a Go/No go tactile discrimination task, where they had to discriminate between two whiskers vertically arranged in the same column of the whisker pad. Mice were rewarded with a drop of water if they licked a spout upon stimulation of the top whisker (B2) whereas they were punished with a 10-second-long timeout if they licked for the bottom whisker (C2). Once mice became expert at the task and performed stably with a high percentage of correct trials over at least 3 consecutive sessions, we switched the task to a Go/No go visual task. In this condition, we replaced the top and bottom whisker stimulations with “top” and “bottom” visual stimuli. These stimuli consisted of black squares on a gray background drifting along the same rostro-caudal direction than the whisker deflections. The screen was

oriented to be centered and parallel to the right retina on the same side as the previously stimulated whiskers. The location of the moving squares along the vertical axis was chosen to roughly match the location of the whiskers within the visual field of the animals (see Methods).

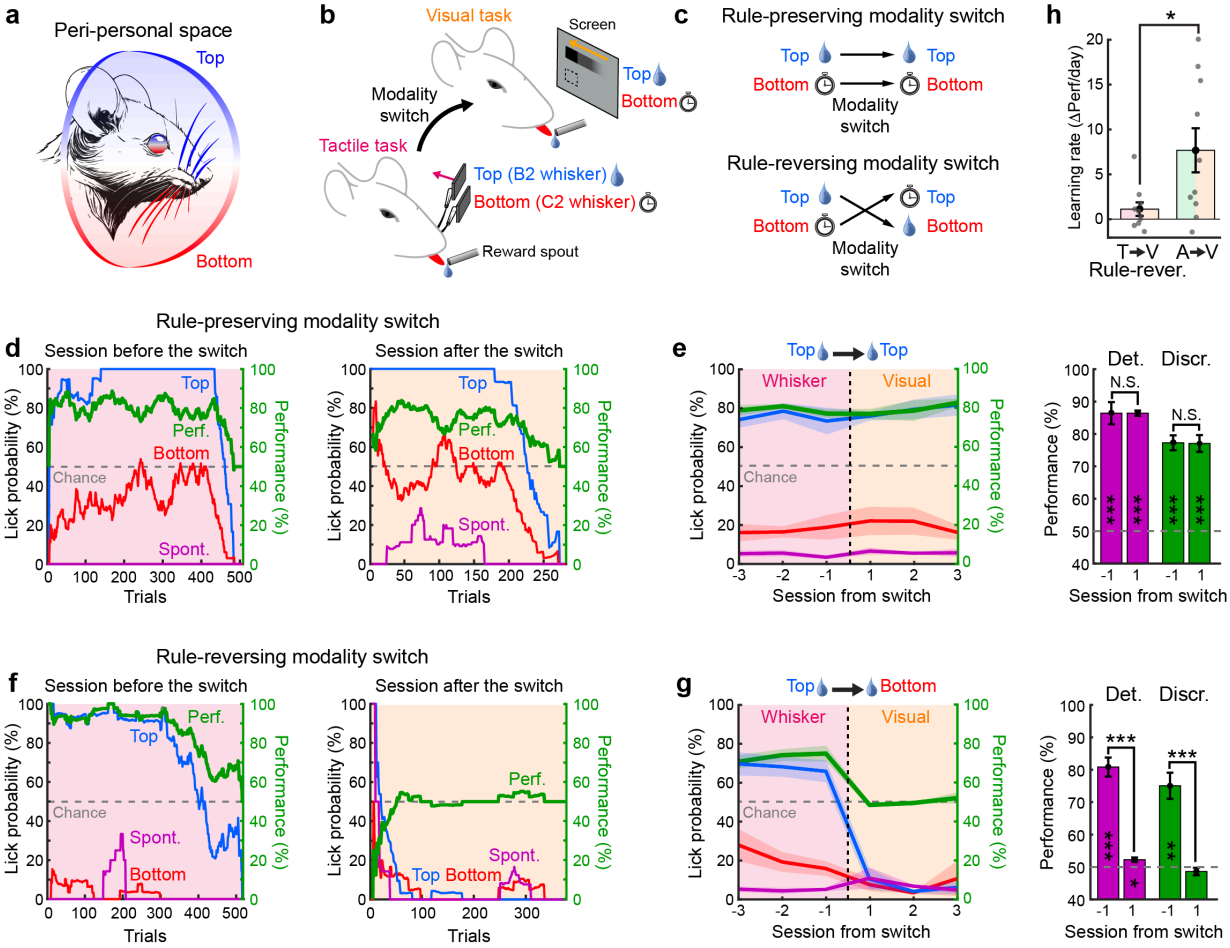


Fig. 1. Cross-modal generalization in mice.

a, Illustration of the common organization of peri-personal space for visual and whisker tactile inputs in mice. **b,** Schematic of the behavioral Go/No go paradigm for studying cross-modal generalization of spatial information in mice. **c**, Two types of tactile-to-visual modality switch: "rule-preserving", wherein the spatial location of rewarded stimuli is preserved, and "rule-reversing", wherein the location of rewarded stimuli is reversed. **d,** Left: example of a session with the tactile task the day before modality switch. Conditional lick probabilities over trials are shown for the top whisker (blue), the bottom whisker (red) and in absence of stimuli (pink). Task performance computed as the percentage of correct discrimination trials (see Methods) is shown in green with chance level as gray dashed line. Traces were computed on a sliding window of 60 trials. Right: same as left for the first visual session following a rule-preserving modality switch. **e,** Left: task performance and conditional lick probabilities averaged across mice (N=5 mice) over consecutive sessions before and after a rule-preserving switch (vertical dashed line). Shaded area: S.E.M. Color code as in **d**. Right: detection (purple) and discrimination (green) performance distribution for the session before and after the switch (two-sided paired t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant).

Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). **f-g,** Same as in **d-e** but for a rule-reversing modality switch. **h,** Comparison of relearning rates between mice that underwent a rule-reversing modality switch and mice that underwent switch from a non-spatial auditory task to the same visual task (N=10 mice for tactile group and N=10 mice for auditory group, unpaired two-sided t-test, *p=0.02).

To test whether mice can use previous associations learned during the tactile task to infer the reward contingency (i.e. the rule that determines when a reward will be delivered following a sensorimotor response) in a new visual task, we considered two scenarios: a “rule-preserving” and a “rule-reversing” modality switch. The cohort of mice undergoing the rule-preserving modality switch could obtain reward by licking for stimuli presented at the same spatial location after the switch. For mice undergoing the rule-reversing switch, the reward contingency was spatially reversed following the switch to the new modality (Fig. 1c). Following rule-preserving modality switches, we observed rapid generalization of the learned association to the new modality. Mice seamlessly performed the task with a comparable level of high performance as observed in the session preceding the switch (Fig. 1d,e). In contrast, task performance was strongly affected following a rule-reversing modality switch, consistently falling at chance level. Mice typically attempted to lick in response to both stimuli during the early part of the session before disengaging and ceasing any licking behavior (Fig. 1f). In some cases, mice attempted to lick first for visual stimuli spatially congruent with the previously rewarded whisker, leading performance to briefly drop below chance level. Following rule-reversing switches, mice displayed a strong resistance to engaging with the task for several consecutive sessions (Fig. 1g). Such behavior persisted despite strong thirst-driven motivation, as indicated by the prolonged licking bouts observed when water drops were manually delivered to maintain task engagement[20]. This implies that the mouse's failure to perform after a rule-reversing

modality switch stems from conflicting prior knowledge of the task rule, rather than from a lack of motivational drive. We verified that this result was not caused by a preference for the top whisker or for the top visual stimulus by repeating the same experiments with cohorts of mice trained to lick for bottom whisker stimulations. Mice maintained high task performance after switches that preserved the spatial rule, regardless of whether they were initially trained on one whisker or another, but performance dropped when the rule was reversed (Extended Data Fig. 1a-c). We further confirmed that mice could also generalize the task rule from the visual to the tactile modality, demonstrating bidirectional cross-modal generalization of learned associations (Extended Data Fig. 1g-l).

Despite the limited visibility of the capillary glass tubes used for whisker stimulations in the dark, we tested whether mice could potentially rely on tubes moving as visual cues to perform the tactile task, thereby generalizing within the visual domain instead of across sensory modalities. We carried out control experiments where mice proficient in the tactile task underwent sessions with whiskers temporarily removed from the tubes, and subsequently reintroduced. We observed that both detection and discrimination performance dropped to chance level immediately after the whiskers were removed from the capillary tubes, but recovered to expert levels once the whiskers were reinserted into the piezo stimulators (Extended Data Fig. 2). This demonstrates that mice were not using visual cues to perform the whisker discrimination task in the dark.

Besides generalizing task rule through common spatial organization, mice could also potentially generalize the abstract Go/No go structure common to the two tasks (i.e. “acting upon one stimulation, the Go stimulus, leads to reward while acting upon the other, the No go stimulus, leads to a timeout”) to rapidly increase

performance after the switch. To investigate this possibility, we trained mice on an auditory Go/No go discrimination task, which produced similar levels of motivation without prior knowledge of spatial rules related to reward contingency. Mice were then switched to the spatial visual task (Extended Data Fig. 1d). The stimuli used for the auditory task were two pure tones played from the same speaker, bearing no clear spatial relationship to the visual stimuli introduced after the modality switch. In absence of spatial prior, performance dropped to chance level after the switch but steadily recovered to expert level over the next few sessions (Extended Data Fig. 1e,f). We concluded that mice could learn the task after modality switch significantly faster in absence of the conflicting spatial prior observed for mice experiencing rule-reversing modality switches (Fig. 1h).

Finally, we compared our results for rule-reversing modality switches with conditions in which reward contingencies were switched within the tactile modality, as previously studied[21,22]. After mice reached expert level in the tactile task, we reversed the reward contingencies between the two whiskers. This led to a stark drop in discrimination performance below chance level, indicating that mice rigidly persisted in performing the task following the original rule (Extended Data Fig. 3). Following the switch, performance increased but remained below or at chance level for at least three consecutive sessions. Thus, mice behave differently when reward contingencies are switched within the same sensory modality as they continue to inflexibly produce the same sensorimotor transformation, likely reflecting ingrained habitual behaviors.

Co-aligned visuo-tactile spatial maps in the dorsal cortex

Our behavioral results suggest that mice can generalize previously learned sensorimotor associations across sensory modalities (Extended Data Fig. 4), based on a common representation of the peri-personal space. Cortical circuits are necessary for conscious perception and are believed to mediate flexible goal-directed behaviors such as cross-modal generalization[2,23]. To pinpoint the cortical regions responsible for visuo-tactile generalization, we mapped the topographic representation of vertical space for both modalities in the dorsal cortex of transgenic mice expressing the calcium indicator GCaMP6f in cortical layers 2/3 (Fig. 2a, see Methods). We first used established retinotopic and somatotopic mapping protocols[20,24] to identify whisker-responding and retinotopically organized cortical areas through a 5 mm diameter cranial window over the posterior part of the left dorsal cortex (Extended Data Fig. 5a). Whisker response patterns and retinotopic sign maps enabled us to precisely fit a projection of the Allen Mouse Brain Atlas to the cranial window of each mouse (see Methods). We used this atlas to register all functional maps into a common reference frame (Extended Data Fig. 5b-f).

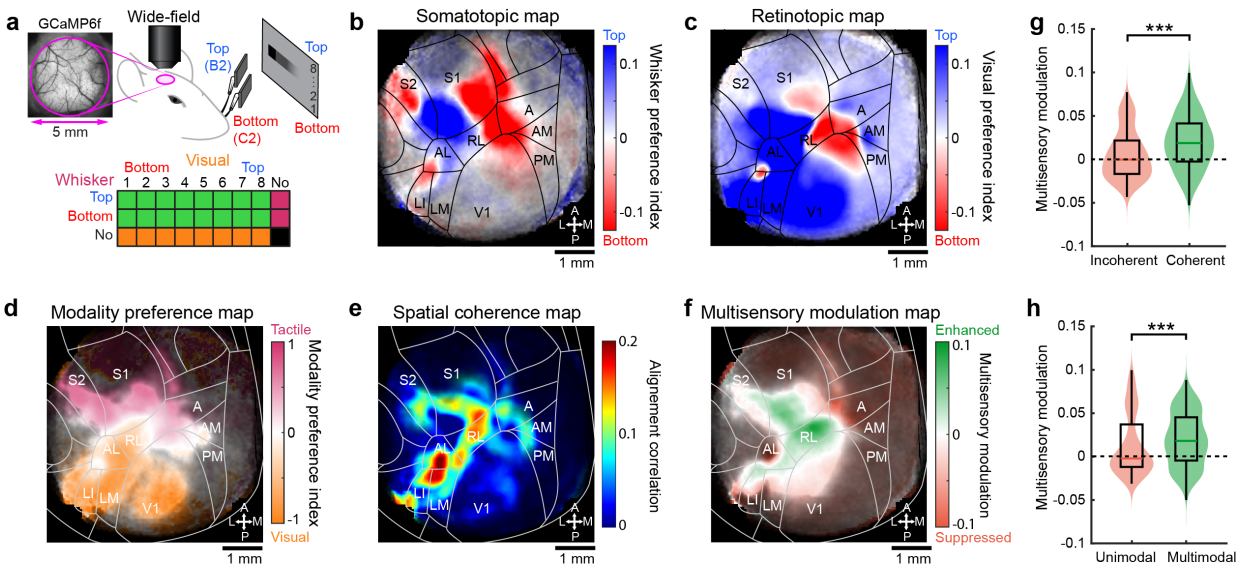


Fig. 2. Co-aligned visual and tactile functional maps in the dorsal cortex.

a, Top: schematic of the visuo-tactile sparse noise protocol used during wide-field imaging of GCaMP6f through a cranial window. Bottom: matrix of all combinations of visual and whisker stimuli (green: multisensory, orange: visual, magenta: tactile). **b,** Somatotopic map of vertical

space computed from whisker stimuli averaged across mice (N=29 mice for all panels of the figure), with transparency defined by response significance in each pixel (see Methods). A projection of the Allen Mouse Brain Atlas is overlaid on top with areas names and orientation. **c,** Average retinotopic map of vertical space. **d,** Average modality preference map between visual and tactile responses. **e,** Average spatial coherence map between visual and tactile representations (see Methods). **f,** Average multisensory modulation map comparing visuo-tactile responses and combination of unisensory responses (see Methods). **g,** Multisensory modulation index for pixels belonging in regions of high spatial coherence compared to regions without spatial coherence (n=26,286 pixels versus n=28,394 pixels, unpaired two-sided Wilcoxon test, ***$p<10^{-300}$). **h,** Multisensory modulation index for pixels belonging to unimodal or multimodal regions (n=10,264 pixels versus n=15,921 pixels, unpaired two-sided t-test, ***$p<10^{-116}$).

To characterize the cortical representation of the vertical space for both unisensory and multisensory stimulations, we designed a visuo-tactile sparse noise protocol (Fig. 2a, see Methods). This protocol was used in task-naïve mice to obtain retinotopic and somatotopic maps for vertical space by computing the preferred spatial position for each pixel. In response to whisker stimuli, we identified the well-established somatotopic arrangement of the primary and secondary whisker somatosensory cortices, S1 and S2, which exhibited a topographic inversion at their boundary (Fig. 2b). Importantly, we observed that whisker stimulations also evoked organized somatotopic maps in several visually responsive areas including the anterior (A), rostro-lateral (RL), antero-lateral (AL) and latero-intermediate (LI) areas. This suggested that whisker representations might be present in a more extended cortical network than previously reported[25]. Strikingly, the maps obtained with visual stimuli displayed a very similar organization to the tactile ones in these associative areas as well as in S1 and S2 (Fig. 2c). The extended spatial representations evoked by visual or tactile stimuli were found consistently across mice (Extended Data Fig. 5g). This suggests that spatially localized stimuli, regardless of their visual or tactile nature, might share a common topographic representation that facilitates mapping between sensory modalities, as previously observed in the superior colliculus[16,17].

Spatial representations evoked by these two modalities displayed an angular offset that we estimated around 30 degrees by comparing the angle difference between gradient vectors obtained from these maps (see Methods). This might reflect the mouse's internal model of how whisker sensations align with their corresponding locations in the visual field[13].

We further investigated the functional properties of these representations by first computing a modality preference index to assess what sensory modality dominates each area (Fig. 2d). As expected, S1 and S2 were dominated by tactile inputs whereas V1 was dominated by visual inputs. RL and the region at the border between AL and LI displayed a more balanced preference for both modalities. In addition, we measured the spatial coherence between maps of vertical retinotopy and somatotopy indicating local co-alignment (Fig. 2e, see Methods). This confirmed a widespread spatial co-alignment across most associative areas in the belt between V1 and S1. We further computed a multisensory modulation index, comparing multisensory responses triggered by visuo-tactile stimuli to the maximal unisensory response on a pixel-by-pixel basis (see Methods). The resulting map revealed a strong multisensory enhancement in visuo-tactile associative areas and in S1 (Fig. 2f). Multisensory responses were comparable under visuo-tactile conditions where whisker stimuli were synchronous or delayed by 0.15 seconds to ensure simultaneity of evoked responses in the cortex (Extended Data Fig. 6), with a tendency to show stronger enhancement in the latter case, as previously documented[19]. Moreover, we found that multisensory enhancement was more pronounced in regions with higher coherence between spatial maps and with strong bimodal representations (Fig. 2g,h).

254 **Anatomical origin of spatial maps in the dorsal cortex**

255 Functional maps measured with wide-field calcium imaging could result from
256 direct inputs from visual and tactile primary cortical areas, from evoked top-down
257 inputs[26], or even be the result of highly stereotypical uninstructed movements evoked
258 by sensory stimuli[27]. To investigate the synaptic origin of these maps, we performed
259 anatomical experiments to map feedforward and feedback projections between
260 primary sensory areas and associative areas displaying visuo-tactile representations
261 (Fig. 3a). We obtained visual and tactile functional maps for representation of vertical
262 space in wild-type mice using intrinsic optical signal imaging under low isoflurane
263 anesthesia (Extended Data Fig. 7a,b,d,e, see Methods). These maps were then used
264 to identify two cortical locations representing distinct iso-horizontal vertical positions
265 in V1 or to target B2 and C2 barrels in S1. We then opened the cranial window and
266 injected two adeno-associated viral vectors to induce expression of tracer proteins
267 GFP and tdTomato in the respective locations (Extended Data Fig. 7c,f). After 10-15
268 days, transcardial perfusions were performed and brains were extracted, flattened,
269 and sliced (see Methods). Enriched M2 subtype muscarinic acetylcholine receptors
270 (M2 AChR) in V1 and S1 barrel field were used as landmarks to fit the Allen Mouse
271 Brain Atlas on the reconstructed stack, confirming the location of injection sites along
272 the vertical representation of V1 and S1 (Fig. 3b,c). Axonal projections from primary
273 sensory areas were found in associative cortical regions where visuo-tactile responses
274 were measured with the same spatial organization. Variability in injection sites allowed
275 us to create an anatomical map that described axonal preferences for top and bottom
276 location (Fig. 3b,c), aligning closely with wide-field imaging results (Fig. 2b,c). This

confirmed that the functional maps are inherited, at least in part, from direct feedforward projections from primary cortical areas.

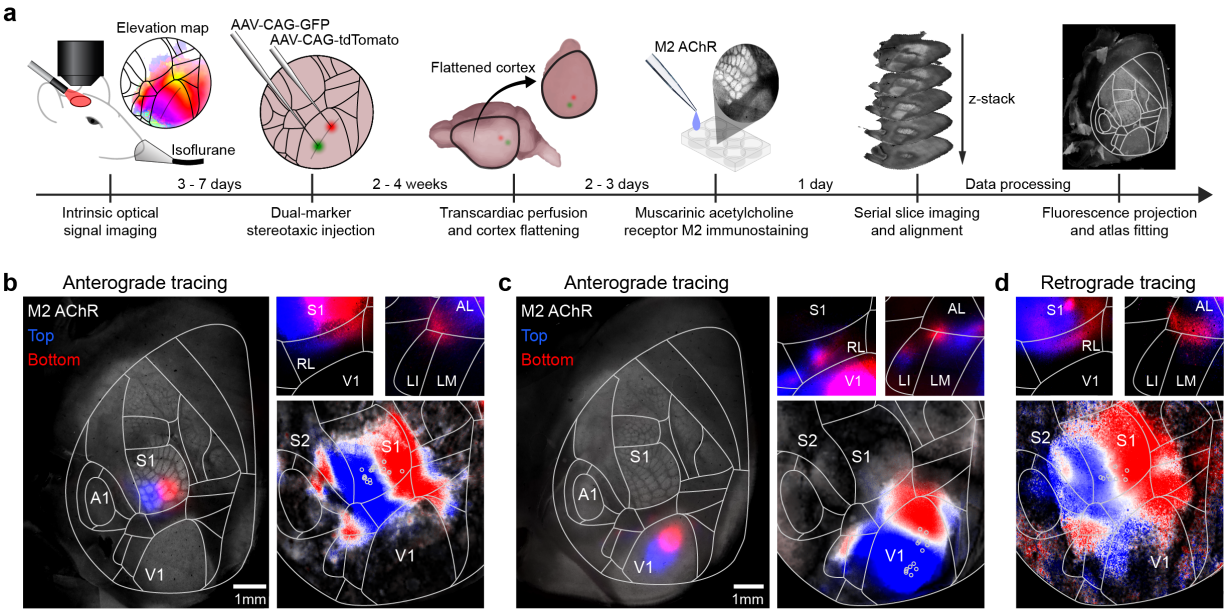


Fig. 3. Co-aligned visual and tactile anatomical projection maps in the dorsal cortex.

a, Schematic of the data collection and analysis pipeline for topographic anatomical tracing. **b,** Anterograde labeling of S1 projections using viral vectors injections. Left: injection sites in B2 and C2 barrels in whisker S1. Right: conserved somatotopic organization of projections in associative areas for the cortex shown in the left (top). Bottom: Reconstructed whisker preference averaged across mice using all injection sites depicted by circles (see Methods, N=9 mice and n=18 injections, Pearson coefficient of correlation between anatomical and wide-field map: 0.8311, $p<10^{-300}$). **c,** Same as **b** but for anterograde labeling of V1 projections with two injection sites along the iso-horizontal axis (N=7 mice and n=14 injections, Pearson coefficient of correlation: 0.5121, $p<10^{-300}$). **d**, Retrograde labeling of S1-projecting neurons using CTB-555 and CTB-647. Top: Examples of CTB-labelled neurons spatially organized in associative areas. Bottom: Reconstructed map of preferred whisker in projecting neurons over the dorsal cortex combining all injection sites (N=5 mice and n=8 injections, Pearson coefficient of correlation: 0.663, $p<10^{-300}$).

While visual stimuli could evoke organized responses in S1 (Fig. 2c), no direct projections were found between V1 and S1[28] suggesting the existence of spatially organized feedback projections from associative areas to S1. Previous work has shown that feedback projections from higher visual areas (including A, RL, AL and LI) to V1 are spatially organized along the vertical dimension[29] but it is not clear if this holds true for feedback projections to S1. Feedback projections from associative areas to S1 were characterized using the same strategy, but with injections of Cholera Toxin

301 Subunit B (CTB) conjugated either with Alexa555 or Alexa 647 (Fig. 3d). S1-projecting
302 neurons for top and bottom locations aligned with wide-field imaging (Fig. 2b) and
303 anterograde tracing (Fig. 3b). We additionally confirmed that the same was true for V1
304 (Extended Data Fig. 7g) as previously reported[29]. Thus, we observed a shared spatial
305 organization of feedforward and feedback projections between primary and
306 associative areas.

307

308 **Single-neuron visuo-tactile functional properties**

309 Spatially organized feedforward and feedback projections could facilitate
310 generalized sensorimotor learning through transfer of spatial information across
311 sensory modalities. However, functional maps obtained with wide-field imaging do not
312 reveal precise computations performed at single-cell level and could still be prone to
313 artifacts produced by neuronal processes originating from other brain structures. To
314 overcome this limitation and assess if neurons in associative areas can mediate cross-
315 modal generalization, we performed two-photon calcium imaging in a subset of mice
316 (Fig. 4a). Single neuron somatic GCaMP6f signal was extracted during the visuo-
317 tactile sparse noise protocol in fields-of-view covering different cortical areas identified
318 with the atlas (Extended Data Fig. 8a-f). Given the observed fluorescence response
319 patterns to various unimodal and multimodal stimuli (Fig. 4b), we could map specific
320 functional properties of single neurons onto the reference atlas, pooling data across
321 mice, for comparison with corresponding wide-field regions. Many recordings were
322 performed across a large portion of the dorsal cortex to extensively cover responsive
323 visuo-tactile areas (Extended Data Fig. 8g-h). We reconstructed the somatotopic map
324 of vertical space across the cranial window, which closely aligned with the wide-field
325 map (Fig. 4c). Neurons with whisker tactile responses were found across the belt of

associative areas following the somatotopic organization. This further confirmed the existence of an extended network of whisker responsive and visuo-tactile cortical regions[25]. The same was true for neurons responding to visual stimuli which were found across most visual and tactile areas, even extending to S1 and S2 (Fig. 4d). Reconstructed population maps based on single-neuron properties consistently matched with functional maps measured with wide-field calcium imaging (Extended Data Fig. 9).

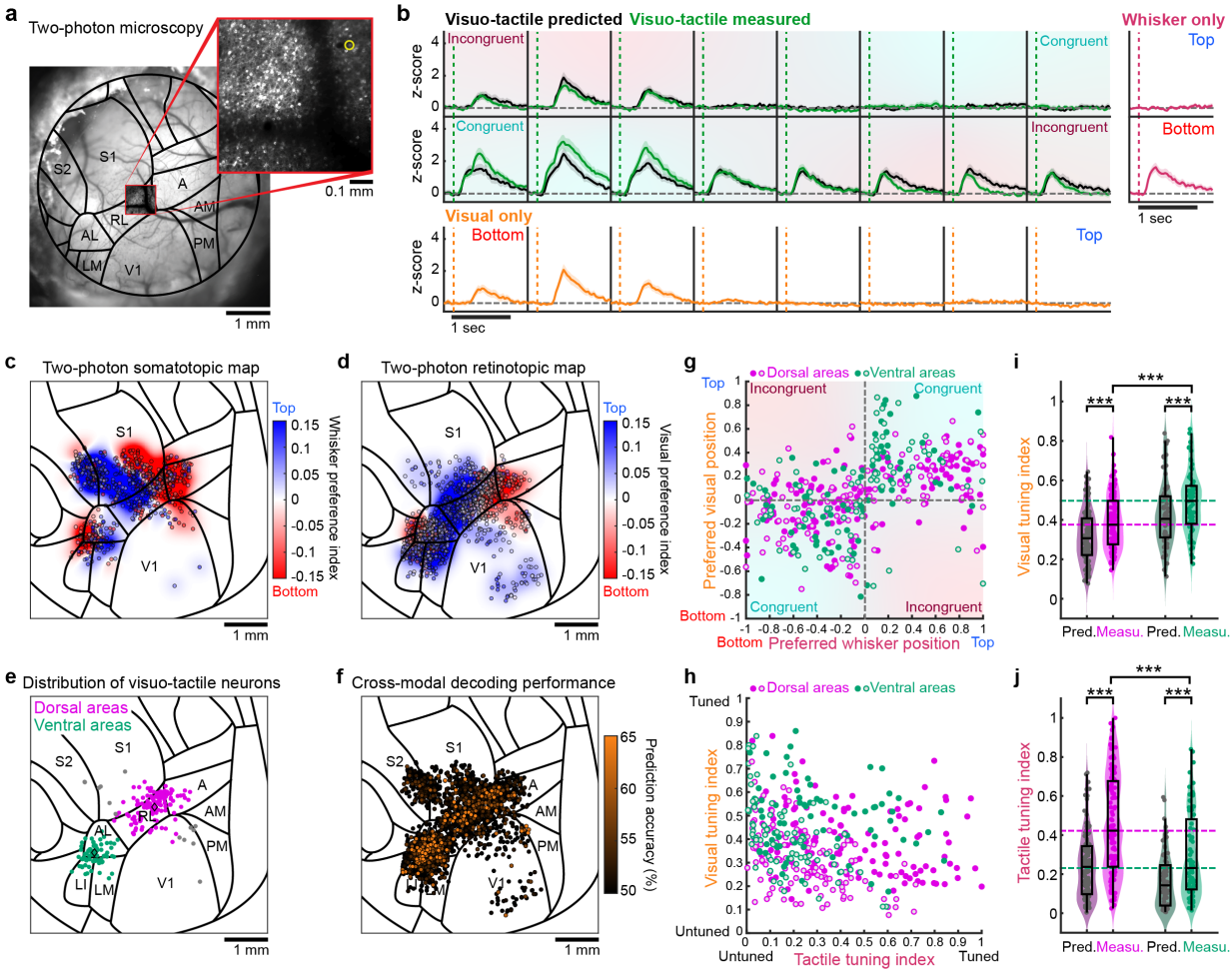


Fig. 4. Visuo-tactile representation of peri-personal space in single neurons.

a, Example cranial window and two-photon field-of-view overlapping with the medial part of area RL. **b,** Response pattern of GCaMP6f fluorescence for the example neuron highlighted with yellow circle in **a**. Unisensory responses (z-scored) are shown on the side for visual (orange) or whisker (magenta) stimuli. Predicted multisensory responses are shown for visuo-tactile stimuli (black) together with measured responses (green). Error bars: S.E.M. **c,** Population of single neurons imaged with two-photon microscopy to measure significant responses to whisker stimuli (n=567 neurons). Each spot represents a single neuron, and the color code indicates preference for top (blue) or bottom (red) stimuli. The reconstructed wide-

field map is displayed in the background (see Methods, Pearson coefficient of correlation: 0.8472, $p<10^{-300}$). **d,** Same as **c** for neurons significantly responding to visual stimuli (n=1,593 neurons, Pearson coefficient of correlation: 0.6505 $p<10^{-300}$). **e,** Distribution of all visuo-tactile bimodal neurons identified. Neurons are classified as being part of the ventral (green) or dorsal (pink) pathway depending on their location (see Methods). Neurons that are not part of these pathways are shown in gray. **f,** Distribution of responsive neurons color-coded by their visual decoding accuracy, following training with whisker-stimulation responses (see Methods, N=2,563 neurons). **g,** Comparison of preferred visual position and preferred whisker along the vertical axis in visuo-tactile stimulations condition for predicted (open circles) and measured (full circles) responses of ventral neurons (n=75 significantly responsive multimodal neurons, Pearson coefficient of correlation: 0.36 for predicted with $p=1.6x10^{-3}$ and 0.48 for measured with $p=1.5x10^{-5}$; 77% of neurons in congruent quadrants) and dorsal neurons (n=124 significantly responsive multimodal neurons, Pearson coefficient of correlation: 0.55 for predicted with $p=3.7x10^{-11}$ and 0.47 for measured with $p=4.3x10^{-8}$; 71% of neurons in congruent quadrants). **h,** Comparison between tactile and visual tuning indices computed from the predicted (open circles) or measured (full circles) visuo-tactile responses of neurons in **e**. **i,** Comparison of the visual tuning indices from **h** between predicted and measured responses for ventral and dorsal stream neurons (two-sided paired Wilcoxon test between measured and predicted for ventral: n=75 neurons, ***$p=1.7x10^{-5}$; for dorsal: n=124 neurons, ***$p=3.1x10^{-11}$; two-sided unpaired t-test comparing dorsal and ventral measured responses: ***$p=3.2x10^{-5}$). **j,** Same as **i** for the tactile tuning indices (two-sided paired Wilcoxon test between measured and predicted for ventral: n=75 neurons, ***$p=7.7x10^{-10}$; for dorsal: n=124 neurons, ***$p=1.6x10^{-19}$; two-sided unpaired t-test comparing dorsal and ventral measured responses: ***$p=1.1x10^{-5}$).

Among all neurons imaged, we identified four functional cell types. Unisensory visual or tactile neurons only responded to their respective modality, gated neurons responded only when both visual and tactile stimuli were presented together, and bimodal visuo-tactile neurons responded to both unisensory visual and whisker stimuli (Extended Data Fig. 10a-b). Importantly, bimodal visuo-tactile neurons localized into two distinct clusters (Fig. 4e), corresponding to the multimodal domains identified by wide-field imaging (Fig. 2d), and consistent with anatomical projection patterns (Fig. 3b,c). As these clusters coincided with areas associated with the dorsal (A, RL) and ventral (AL, LI) streams, we used these designations moving forward. Due to their distinct functional properties, visuo-tactile neurons may facilitate cross-modal generalization. A Bayesian decoder trained to discriminate whisker stimuli based on individual neuronal responses was evaluated for generalization to visual stimuli (see

379 Methods). Visuo-tactile neurons in associative areas uniquely enabled effective
380 generalization across sensory modalities (see Fig. 4f). Other neuron types
381 demonstrated significantly lower decoding accuracy when we tested their ability to
382 generalize the tactile discrimination to the visual modality, particularly with larger
383 populations (see Extended Data Fig. 10c,d).

384 We further characterized the response properties of single neurons for visuo-
385 tactile stimuli in comparison with their responses predicted from unisensory responses
386 (Fig. 4b, see Methods). We found that bimodal neurons were typically tuned to
387 spatially congruent stimuli across both modalities. The example neuron in Figure 4b
388 responded preferentially to the bottom part of the visual field and to the bottom whisker.
389 Using unisensory responses, we predicted the response pattern to visuo-tactile stimuli
390 as the maximum response between the two modalities for each combination. When
391 comparing the predicted response with the measured one, we observed suppression
392 in incongruent combinations and enhancement in congruent ones. This nonlinear
393 modulation profile enables neurons to maintain their spatial selectivity independently
394 of the stimulated modality (visual, tactile, or visuo-tactile). This response property is
395 reminiscent of the supramodal encoding of object orientation reported in visuo-tactile
396 neurons of the rat posterior parietal cortex[6]. Population analysis confirmed that spatial
397 congruence in multisensory selectivity is prevalent among populations of multimodal
398 neurons in both the ventral and dorsal stream (Fig. 4g). Additionally, the two neuronal
399 populations exhibited sharper tuning for whisker and visual positions than that
400 predicted by their unisensory responses alone (Fig. 4h). Neurons from the ventral
401 domain displayed sharper tuning for particular visual locations, whereas neurons in
402 the dorsal domain exhibited greater tuning to individual whiskers (Fig. 4i,j), consistent
403 with modality preferences observed in wide-field imaging data (Fig. 2d). These results

404 showed that bimodal neurons in visuo-tactile associative areas are indeed specifically
405 responding to spatially congruent visuo-tactile stimuli and could mediate cross-modal
406 generalization during goal-directed tasks.

407

408 **Cross-modal generalization loss-of-function**

409 Cross-modal decodability of stimulus location observed at single-cell level (Fig.
410 4f), alongside the anatomical connectivity between visual and whisker somatosensory
411 cortices (Fig. 3), suggest that associative visuo-tactile cortical areas could mediate
412 cross-modal generalization of goal-directed behaviors. We performed loss-of-function
413 experiments to assess the necessity of these areas for visuo-tactile generalization.
414 Given that task-naïve mice (those not yet exposed to any behavioral tasks) exhibited
415 co-aligned and anatomically connected spatial maps, we reasoned that the
416 reverberation of evoked responses across this extensive visuo-tactile network could
417 potentially drive learning processes beyond the initially stimulated modality. This could
418 facilitate cross-modal generalization before any direct exposure to the second task
419 with the other modality. To prevent this possibility, we decided to chronically silence
420 associative cortical areas prior to any sensorimotor learning. We virally expressed the
421 tetanus toxin light chain (TeNT-P2A-GFP) in ventral and dorsal areas, thereby
422 preventing neurotransmitter vesicle release in transfected neurons (Fig. 5a, see
423 Methods). Neurons expressing TeNT also co-expressed GFP, allowing comparison of
424 the expression pattern with the fitted atlas (Fig. 5b). After subtracting blood vessel
425 patterns and comparing with the cranial window before injection (see Methods), we
426 characterized the extent of GFP expression and overlap with different visuo-tactile
427 areas (Extended Data Fig. 11a-d). To ensure strong expression of TeNT to effectively
428 suppress vesicle release in the transfected neurons before beginning the behavioral

429 training, we waited at least four weeks after viral injections[30]. Mice normally learned
430 the whisker discrimination task and reached expert performances comparable to those
431 observed in mice trained without TeNT expression (Fig. 5c, Extended Data Fig. 11a-
432 d). However, when switching to the visual task in the rule-preserving condition,
433 performance typically dropped to chance level (Fig. 5d,k).

434 To identify which visuo-tactile areas are necessary for cross-modal
435 generalization, we expressed TeNT-P2A-GFP in different cortical locations across
436 mice (see Methods). We were able to cover all visuo-tactile areas with varying degrees
437 of overlap with the expression of TeNT (Extended Data Fig. 11e). Taking advantage
438 of this variability, we calculated for each cortical area the extent to which the overlap
439 of its surface with TeNT expression correlated with any decline in performance
440 following a rule-preserving switch (Extended Data Fig. 11f). This analysis revealed
441 that generalization impairment was significantly correlated with TeNT expression in
442 area RL only (Fig. 5e). To confirm this result, we also performed a reverse correlation
443 analysis mapping the average TeNT coverage that evoked a complete impairment of
444 cross-modal generalization (see Methods). Here again, RL silencing was found to
445 consistently prevent cross-modal generalization, revealing that this associative area
446 is necessary for the transfer to occur (Fig. 5f). Grouping mice depending on whether
447 TeNT expression pattern was located in the dorsal or ventral stream (Fig. 5g,i), we
448 confirmed that dorsal stream silencing severely affected generalization performance
449 (Fig. 5h), leading to slower re-learning after rule-preserving switches.

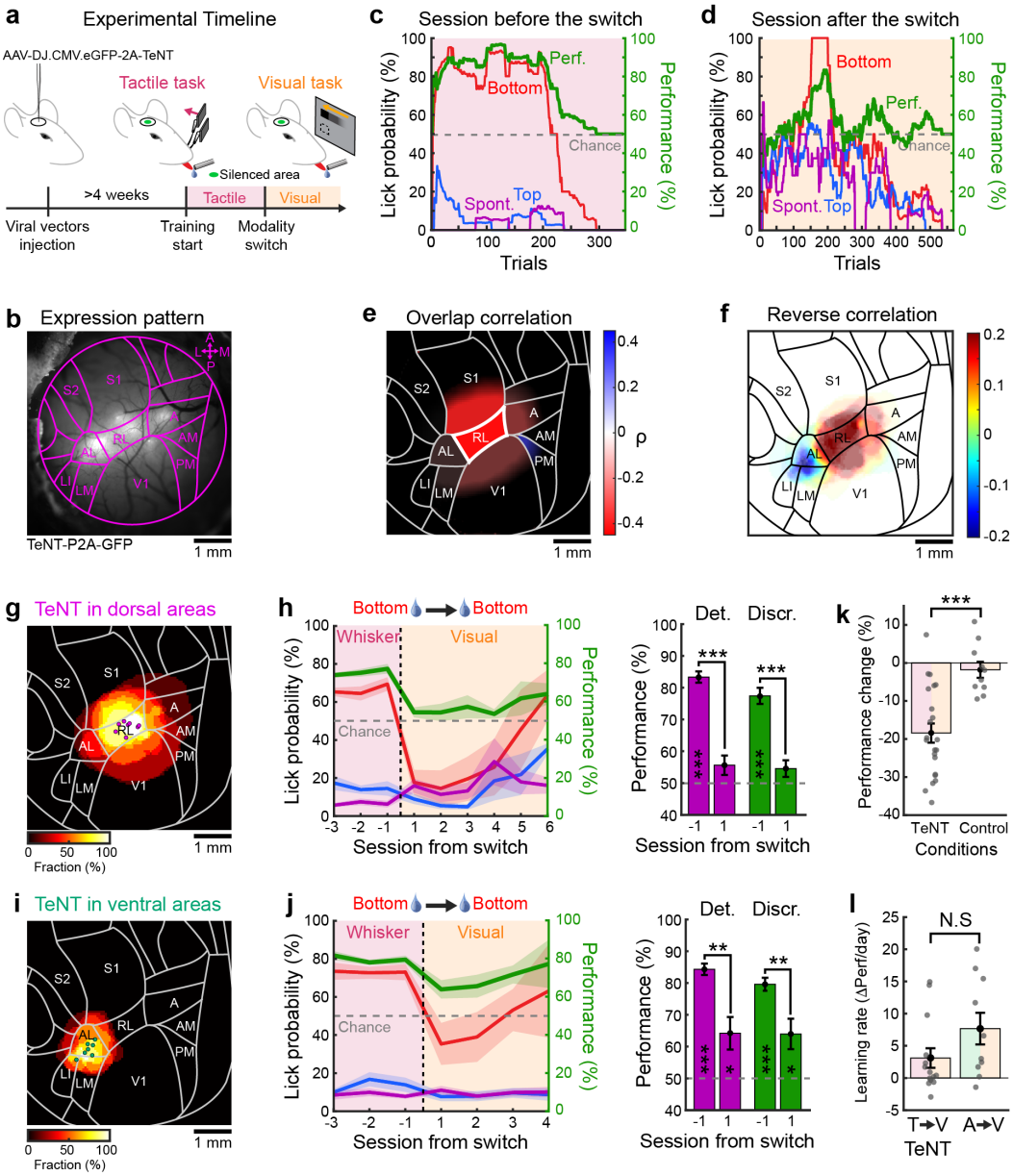


Fig. 5. Area RL is necessary for cross-modal generalization.

a, Schematic of the experimental timeline for loss-of-function experiments. **b,** Cranial window with expression of TeNT-P2A-GFP where viral vectors were injected. Atlas overlaid for reference. **c,** Example session before a rule-preserving modality switch from a tactile to a visual task. Color code as in Fig. 1d. **d,** Session following the modality switch. **e,** Area-based correlation between TeNT-P2A-GFP expression overlap and performance drop following modality switch. Color map indicates Pearson coefficients of correlation ρ. Areas with $p<0.05$ are indicated with a thick border (RL with $p=0.047$). **f,** Average TeNT-P2A-GFP coverage for mice with impaired cross-modal generalization (see Methods). The map is displayed after subtraction of the average coverage across all injected mice. **g,** Average TeNT-P2A-GFP coverage of all mice where only dorsal neurons were silenced (N=8 mice). Dots indicate center-of-mass location for each mouse. **h,** Left: Average task performance and conditional lick probabilities across sessions for mice in **g** with rule-preserving modality switch (vertical dashed line). Shaded area: S.E.M. Color code as in **c**. Right: Detection (purple) and discrimination (green) performance distribution the day before and after the switch (paired two-sided t-test comparing days, *$p<0.05$, **$p<0.01$, ***$p<0.001$, N.S. Not significant).

Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). **i,** Same as **g** but for TeNT-P2A-GFP expression in the ventral stream (N=7 mice). **j**, Same as **h** for ventral areas (paired two-sided t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). **k,** Comparison of performance change following rule-preserving switch between all mice expressing TeNT-P2A-GFP and control mice described in Fig. 1e and Extended Data Fig. 1b (N=22 mice for TeNT and N=10 mice for control, two-sided unpaired t-test, ***p=2.6×10^{-4}). **l,** Learning rate estimated over the first three sessions following modality switch for mice expressing TeNT-P2A-GFP with impaired cross-modal generalization and control mice trained to the auditory task first described in Extended Data Fig. 1e-f (N=14 mice for TeNT and N=10 mice for control, two-sided unpaired t-test, N.S. p=0.11).

In contrast, silencing areas in the ventral stream did not prevent most mice from performing above chance level following a rule-preserving switch (Fig. 5j). Although mice expressing TeNT in RL initially failed to detect or discriminate visual stimuli after task switch, they regained performance in the visual task after several training sessions (Fig. 5h). This indicates that despite a comparable capacity to perform the visual or tactile task, these mice were not able to generalize learning across sensory modalities. This suggests a very selective effect of RL silencing in impairing cross-modal generalization abilities but not unisensory skills.

We further assessed how silencing area RL with TeNT affects task performance following rule-reversing modality switch (Extended Data Fig. 12a-b). Following the switch, their performance fell to chance level but recovered faster than what we observed for control mice (Extended Data Fig. 12c). To fully characterize the learning rate following modality switches in TeNT-expressing mice, we focused on all mice that exhibited complete generalization impairment (i.e. displaying performance near chance level on the first session following the switch regardless of whether the condition was rule-reversing or rule-preserving). For these mice, we quantified the re-learning rate over several sessions following the modality switch and observed that the distribution was comparable to that seen during non-spatial auditory-to-visual

497 switches (Fig. 5l). This implies that silencing area RL prevents mice from applying prior
498 knowledge about the spatial task rule, whether it is conflicting or congruent with the
499 new task rule, resulting in the mice learning the visual task without prior spatial
500 knowledge.

501

502 **Optogenetic sensory substitution for cross-modal generalization**

503 To determine whether cortical area RL is not only necessary, but also sufficient
504 for visuo-tactile generalization, we conducted gain-of-function experiments by
505 substituting visual stimuli after modality switch with direct optogenetic stimulations of
506 RL "top" and "bottom" encoding subregions (Fig. 6a). Projections from primary sensory
507 areas to the ventral and dorsal visuo-tactile stream being largely distinct (Extended
508 Data Fig. 7h-j), we assumed that optogenetic activations of RL would remain restricted
509 mostly to the dorsal stream. Functional mapping was first performed to identify
510 domains of RL encoding top and bottom visual stimuli in Ai32 transgenic mice
511 expressing Cre-dependent ChR2-eYFP. Viral vectors AAV1.CaMKIIa.Cre were then
512 injected in this area to broadly express ChR2-eYFP in RL (Fig. 6b). We employed a
513 projector-based microscope to precisely shape blue light patterns to selectively excite
514 subregions of RL encoding top and bottom stimuli. This approach allowed us to
515 substitute visual stimuli in the upper or lower visual field with precise spatio-temporal
516 optostimulations following the modality switch (see Methods).

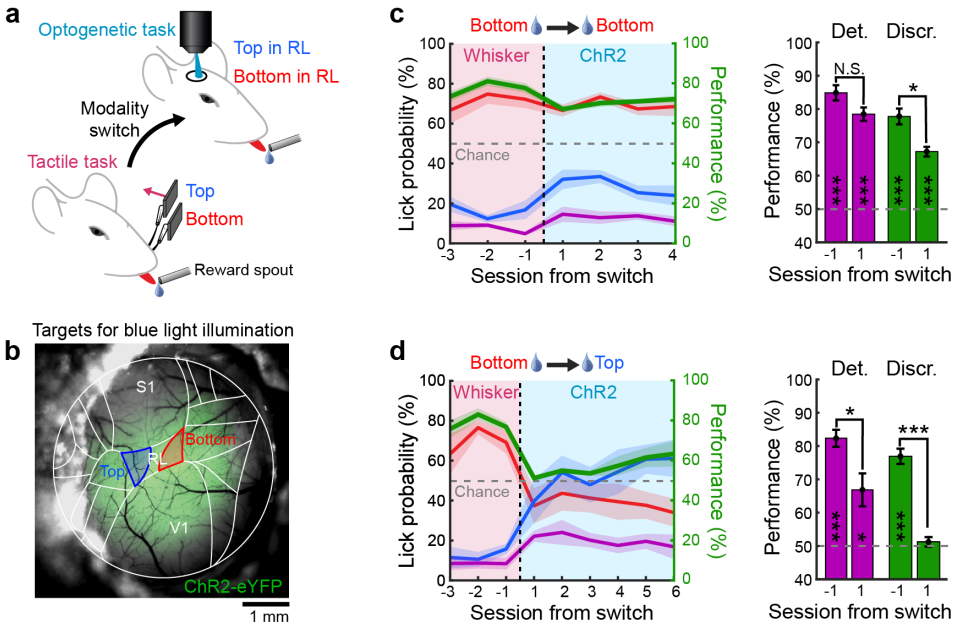


Fig. 6. Area RL is sufficient for cross-modal generalization.

a, Schematic of the behavioral experiment with direct ChR2-mediated optostimulations of RL subregions following modality switch. **b,** Example cranial window of an Ai32 mouse expressing ChR2-eYFP after injection of AAV1.CaMKIIa.Cre viral vector in area RL. Atlas overlaid for reference. Subregions of RL encoding top or bottom stimuli are indicated in blue and red, respectively. Blue light patterns are shaped to match these subregions (see Methods). **c,** Left: Average task performance and conditional lick probabilities across sessions for mice experiencing a rule-preserving modality switch from the whisker task to the optogenetic task (N=7 mice). Right: detection (purple) and discrimination (green) performance distribution the session before and after the switch (paired two-sided t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). **d,** Same as **c** but for mice undergoing a rule-reversing switch (N=6 mice, paired two-sided t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant).

After mice began responding to optogenetic blue light stimulations, we assessed their capacity to transfer previously learned sensorimotor associations from the whisker task to the optogenetic task, both when the spatial rule was preserved or reversed. Following rule-preserving modality switch, mice maintain a comparable level of discrimination performance (Fig. 6c). In contrast, mice undergoing the rule-reversing modality switch performed at chance level after the switch and gradually relearned the task in the following sessions (Fig. 6d). This suggests that activity in RL alone is sufficient to induce cross-modal generalization.

540 **Network model for generalized sensorimotor learning**

541 Based on our experimental findings on the functional and anatomical
542 organization of visuo-tactile cortical circuits, as well as the effect of RL silencing on
543 cross-modal generalization, we developed a network model to recapitulate the
544 behavior we observed in mice. Our model included areas V1, S1, and RL, each
545 comprising two recurrent networks of excitatory and inhibitory neurons encoding "top"
546 and "bottom" spatial location, respectively. Each of these areas projected to reward-
547 modulated decision variables representing a downstream motor area (Fig. 7a).
548 Training the model on the whisker discrimination task involved external inputs to S1
549 (representing thalamic inputs) and reward-modulated learning via Hebbian plasticity
550 of the sensorimotor connections (see Methods). Model simulations showed that after
551 switching inputs to V1, mimicking a rule-preserving switch to the visual modality,
552 performance remained stable (Fig. 7b). However, if the spatial rule of the task was
553 reversed upon modality switch, performance dropped below chance level and then
554 slowly recovered (Fig. 7c, Extended Data Fig. 13). This delay in performance recovery
555 was due to the model weights needing to override the previously learned structure
556 before learning the new association. Mimicking the loss-of-function experiments, we
557 silenced the area RL in the model during training, causing the neural network to learn
558 the visual task from scratch after the rule-preserving switch without any spatial prior in
559 the weight structure (Fig. 7d). The absence of direct connections between the modeled
560 sensory areas prevented ongoing learning in the non-stimulated area. The model
561 suggests that the activity reverberation mediated by bidirectional connections induces
562 a consistent weight structure across connected areas. These weights, however,

needed to be reorganized following a rule-reversing modality switch, leading to delayed relearning time (Extended Data Fig. 13c-d).

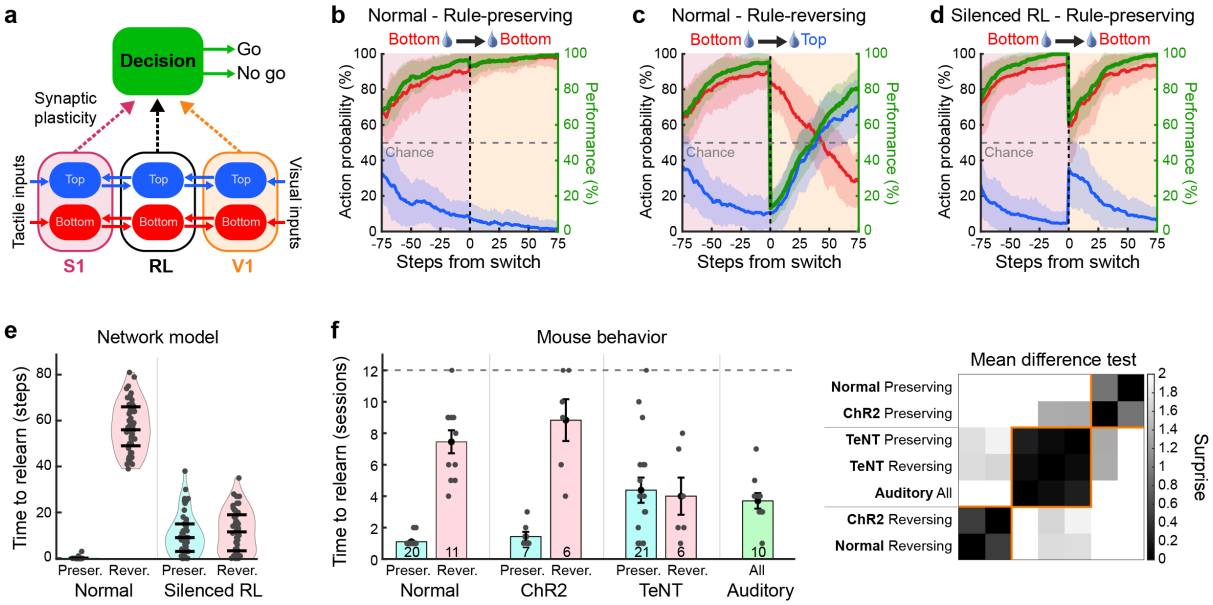


Fig.7. Neural network architecture for cross-modal generalization.

a, Schematic of the neural network model for cross-modal generalization. Synapses projecting to the decision-computing area undergo synaptic plasticity during sensorimotor learning. **b,** Action probability conditional on inputs to S1 before and inputs to V1 after, a rule-preserving modality switch (vertical dashed line). Red: rewarded bottom stimulations, blue: non-rewarded top stimulations. Discrimination performance is shown in green. Shaded areas: standard deviation with n=50 simulations. **c,** Same as **b** after a rule-reversing modality switch. **d,** Same as **b** with silenced RL. **e,** Number of steps necessary to cross a performance threshold of 75% in the full network model (left) or in the model with silenced RL (right) after rule-preserving ("Preser.") or rule-reversing ("Rever.") modality switches (n=50 simulations in all except "Normal Rever." for which n=49). Error bars indicate median, first and third quartiles. **f,** Left: Number of sessions needed to cross a performance threshold of 65% following the modality switch. Training was stopped after 12 sessions (dashed line) and mice that did not reach the criteria before this session are plotted on the dashed line. Mice numbers are indicated for each group at the bottom of the bar. Error bars: S.E.M. Right: Surprise matrix computed from pairwise unpaired two-sided t-test between conditions. A hierarchical clustering based on cosine similarity was used to group conditions based on surprise values.

To quantify the effect of different task conditions in the learning trajectory following modality switch in the model, we computed the time needed in simulation steps to recover expert level. Following a rule-preserving modality switch, the time to relearn was near-instantaneous whereas it was much longer following rule-reversing switches (Fig. 7e). Interestingly, when RL was silenced, the time to relearn was in-

588 between these two distributions reflecting that the model learned the task from scratch
589 (Fig. 7e). We compared these results to the time needed to relearn the task after
590 modality switch in all conditions used with our mice (Fig. 7f). In line with the model,
591 mice rapidly recovered high-performance levels after a rule-preserving modality switch.
592 In contrast, time to relearn the task was dramatically prolonged following rule-
593 reversing switches. This was true for both visual stimuli and optostimulations following
594 the switch. In TeNT-expressing mice, however, the time to relearn the task was
595 moderate and similar for both rule-preserving and rule-reversing modality switches.
596 Importantly, the time to relearn in these mice was comparable to the time needed to
597 reach expert level in the auditory task in absence of any spatial prior, reminiscent of
598 the "RL silenced" case in the model. The distribution of peak performance observed
599 for our mice was mostly determined by the nature of the sensory inputs after the switch
600 rather than the switch type (Extended Data Fig. 14).

601

602 Discussion

603 Our results elucidate the cortical architecture underlying the abstract
604 representation of peri-personal space. We identified the specific cortical circuit that
605 enables rapid generalization of sensorimotor associations learned from one modality
606 to another, forming the neural substrate for a key component of flexible behavior.

607 Cross-modal generalization has been reported across species including apes[3],
608 rodents[2] and, more recently, bumblebees[4]. Early research into the mammalian brain
609 circuits involved in cross-modal generalization underscores the necessity of different
610 brain structures to process different types of information such as spatial, temporal, and
611 object-related features[23,31,32]. The posterior parietal cortex (PPC) plays an important

role in processing spatial information[33,34]. Our results point out that area RL, a subregion of PPC, is crucial for cross-modal generalization based on abstract representations of peri-personal space. Area RL displays visuo-tactile neuronal responses[19] and is implicated in whisking movements control[35]. It is strategically positioned between the primary visual cortex and the primary whisker somatosensory cortex, facilitating bidirectional information transfer between these sensory systems. Area RL possesses a retinotopic map biased towards the lower part of the visual field[24] where most whiskers are visible, and its neurons are tuned to high binocular disparity, aligning with objects in close proximity, potentially within whiskers' reach[36]. It was also implicated in optic flow processing[37] while its contribution to high-order motion computations remains debated[38,39]. Our findings unveil a fundamental functional specialization of area RL, confirming its unique role in facilitating visuo-tactile coordination and sensory abstraction in the peri-personal space. Future research could elucidate how this area may contribute to sensorimotor integration during active multimodal exploration of nearby objects, similar to multimodal limb posture encoding observed in the primate PPC[40].

Other forms of cross-modal generalization, such as object recognition, may engage different cortical regions. Indeed, areas of the dorsal stream are classically associated with spatial, attentional, and motion processing or motor guidance, while areas of the ventral stream are involved in object recognition. Previous studies have reported visuo-tactile multimodal responses in both the ventral and dorsal pathway in primates[41–43], including areas initially thought to be exclusive to visual processing like V4[44] and MT/V5[45]. Our study uncovered visuo-tactile areas belonging to the ventral or dorsal stream in mice[46,47]. Specifically, RL and A are typically associated to the dorsal stream while LI, with its significant projection to the postrhinal (POR) cortex, is

637 associated with the ventral stream[46]. Lesions in the perirhinal cortex in rats,
638 downstream to POR, obstruct cross-modal object recognition during spontaneous
639 exploration[31]. Moreover, areas LI and AL have been functionally characterized as
640 specialized for shape processing and object recognition[48,49], suggesting that they
641 could represent a rodent homologue of the visuo-haptic subregion of the lateral
642 occipital complex (LOtv) reported in human fMRI studies[41]. These reports suggest that
643 the ventral pathway might contribute specifically to cross-modal object recognition.
644 Future research focusing on objects discrimination could provide further insight into
645 the specific role of visuo-tactile areas in the ventral stream.

646 Wide-field imaging and anatomical mapping experiments revealed a more
647 extensive and organized network of visuo-tactile associative cortical areas than
648 previously identified[19,46,50]. Many associative areas exhibited a shared representation
649 of vertical space between visual and whisker tactile inputs. These spatial maps
650 extended even to primary sensory cortices through associative areas, with strong
651 responses observed in S1 following visual stimulations despite no direct projections
652 between V1 and S1[28]. Although recent findings advise caution in interpreting cross-
653 modal signals between primary cortical areas due to potential confusion with signals
654 evoked by uninstructed movements[27], our results suggest that certain fundamental
655 features, such as spatial location, indeed have a shared representations across
656 various cortical areas. These representations possibly support object or event-
657 oriented abstract encoding[51] for cross-modal generalized learning. Organized
658 feedback projections from associative areas to primary sensory cortices appear to play
659 an unsuspected role for learning, in addition to their known role in shaping functional
660 properties with contextual information[52]. This raises the question of whether the same
661 or different feedback projections subserve these functions.

Consistent with the literature on multisensory integration[53], we observed multisensory modulations that enhanced responses to spatially congruent visuo-tactile stimuli and suppressed responses to incongruent stimuli. This response pattern could potentially arise from surround suppression for incongruent inputs, mediated by local parvalbumin-positive interneurons as observed for conflicting visuo-auditory stimuli[12]. By comparing congruent to incongruent combinations, we uncovered multisensory modulation rules that shape a supramodal representation of peri-personal space. Supramodal encoding of gratings orientation was reported in neurons of the rat posterior parietal cortex (PPC) in response to visuo-tactile stimuli[6] despite a lack of organized functional maps for orientation selectivity in rodents. Further investigations at the single-neuron level could help elucidate the mechanisms of synaptic plasticity underlying the emergence of these supramodal representations during development.

Information generalization across modalities is crucial not only for biological brains but also for artificial systems. With the advent of multimodal large language models such as GPT-4o and Gemini, it is widely recognized that multimodality is fundamental for general artificial intelligence[55,56] and can lead to the emergence of abstract conceptual representations in machine learning systems[57]. We believe that the supramodal encoding of peri-personal space by multisensory neurons we reported could represent an early instance of the same computational mechanism leading to the emergence of conceptual representations[9] that form the building blocks of flexible mental representations known as cognitive maps[1]. In this light, our work represents a major step toward understanding the circuit architecture of one of the most fundamental cognitive computations: sensory abstraction.

687 # References

688 1. Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible
689 Behavior. *Neuron* **100**, 490–509 (2018).

690 2. Cloke, J. M., Jacklin, D. L. & Winters, B. D. The neural bases of crossmodal object
691 recognition in non-human primates and rodents: A review. *Behavioural Brain*
692 *Research* **285**, 118–130 (2015).

693 3. Davenport, R. K. & Rogers, C. M. Intermodal Equivalence of Stimuli in Apes.
694 *Science (1979)* **168**, 279–280 (1970).

695 4. Solvi, C., Gutierrez Al-Khudhairy, S. & Chittka, L. Bumble bees display cross-modal
696 object recognition between visual and tactile senses. *Science (1979)* **367**, 910–
697 912 (2020).

698 5. Bruck, J. N. & Pack, A. A. Understanding across the senses: cross-modal studies of
699 cognition in cetaceans. *Anim Cogn* **25**, 1059–1075 (2022).

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

703 7. Samborska, V., Butler, J. L., Walton, M. E., Behrens, T. E. J. & Akam, T.
704 Complementary task representations in hippocampus and prefrontal cortex for
705 generalizing the structure of problems. *Nat Neurosci* **25**, 1314–1326 (2022).

706 8. Miller, E. K., Nieder, A., Freedman, D. J. & Wallis, J. D. Neural correlates of
707 categories and concepts. *Curr Opin Neurobiol* **13**, 198–203 (2003).

9. Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. & Fried, I. Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005).

10. Garner, A. R. & Keller, G. B. A cortical circuit for audio-visual predictions. *Nat Neurosci* **25**, 98–105 (2022).

11. Godenzini, L. *et al.* Auditory input enhances somatosensory encoding and tactile goal-directed behavior. *Nat Commun* **12**, (2021).

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

13. Weiler, S. *et al.* A primary sensory cortical interareal feedforward inhibitory circuit for tacto-visual integration. *Nat Commun* **15**, 3081 (2024).

14. Jenkinson, E. W. & Glickstein, M. Whiskers, barrels, and cortical efferent pathways in gap crossing by rats. *J Neurophysiol* **84**, 1781–1789 (2000).

15. Sofroniew, N. J., Cohen, J. D., Lee, A. K. & Svoboda, K. Natural whisker-guided behavior by head-fixed mice in tactile virtual reality. *Journal of Neuroscience* **34**, 9537–9550 (2014).

16. Drager, U. C. & Hubel, D. H. Physiology of visual cells in mouse superior colliculus and correlation with somatosensory and auditory input. *Nature* **253**, 203–204 (1975).

17. Drager, U. C. & Hubel, D. H. Topography of visual and somatosensory projections to mouse superior colliculus. *J Neurophysiol* **39**, 91–101 (1976).

729 18. Stein, B. E., Stanford, T. R. & Rowland, B. A. Development of multisensory
730 integration from the perspective of the individual neuron. *Nature Reviews*
731 *Neuroscience* vol. 15 520–535 Preprint at https://doi.org/10.1038/nrn3742
732 (2014).

733 19. Olcese, U., Iurilli, G. & Medini, P. Cellular and synaptic architecture of
734 multisensory integration in the mouse neocortex. *Neuron* **79**, 579–593 (2013).

735 20. Matteucci, G. *et al.* Cortical sensory processing across motivational states during
736 goal-directed behavior. *Neuron* **110**, 4176-4193.e10 (2022).

737 21. Chéreau, R. *et al.* Dynamic perceptual feature selectivity in primary
738 somatosensory cortex upon reversal learning. *Nat Commun* **11**, 3245 (2020).

739 22. Banerjee, A. *et al.* Value-guided remapping of sensory cortex by lateral
740 orbitofrontal cortex. *Nature* **585**, 245–250 (2020).

741 23. Tees, R. The effects of posterior parietal and posterior temporal cortical lesions
742 on multimodal spatial and nonspatial competencies in rats. *Behavioural Brain*
743 *Research* **106**, 55–73 (1999).

744 24. Zhuang, J. *et al.* An extended retinotopic map of mouse cortex. *Elife* **6**, 1–29
745 (2017).

746 25. Yamashita, T. *et al.* Diverse Long-Range Axonal Projections of Excitatory Layer
747 2/3 Neurons in Mouse Barrel Cortex. *Front Neuroanat* **12**, (2018).

748 26. Choi, I., Lee, J.-Y. & Lee, S.-H. Bottom-up and top-down modulation of
749 multisensory integration. *Curr Opin Neurobiol* **52**, 115–122 (2018).

750 27. Bimbard, C. *et al.* Behavioral origin of sound-evoked activity in mouse visual
751 cortex. *Nat Neurosci* **26**, 251–258 (2023).

752 28. Wang, Q. & Burkhalter, A. Area map of mouse visual cortex. *Journal of*
753 *Comparative Neurology* **502**, 339–357 (2007).

754 29. Morimoto, M. M., Uchishiba, E. & Saleem, A. B. Organization of feedback
755 projections to mouse primary visual cortex. *iScience* **24**, (2021).

756 30. Zingg, B., Peng, B., Huang, J., Tao, H. W. & Zhang, L. I. Synaptic Specificity and
757 Application of Anterograde Transsynaptic AAV for Probing Neural Circuitry. *The*
758 *Journal of Neuroscience* **40**, 3250–3267 (2020).

759 31. Winters, B. D. & Reid, J. M. A Distributed Cortical Representation Underlies
760 Crossmodal Object Recognition in Rats. *The Journal of Neuroscience* **30**, 6253–
761 6261 (2010).

762 32. Yeterian, E. H. The effects of visual or auditory cortical lesions on specific cross-
763 modal transfer in the rat. *Neuropsychologia* **15**, 517–527 (1977).

764 33. Whitlock, J. R. Posterior parietal cortex. *Current Biology* **27**, R691–R695 (2017).

765 34. Pasalar, S., Ro, T. & Beauchamp, M. S. TMS of posterior parietal cortex disrupts
766 visual tactile multisensory integration. *European Journal of Neuroscience* **31**,
767 1783–1790 (2010).

768 35. Mohan, H. *et al.* Functional Architecture and Encoding of Tactile Sensorimotor
769 Behavior in Rat Posterior Parietal Cortex. *The Journal of Neuroscience* **39**, 7332–
770 7343 (2019).

771 36. La Chioma, A., Bonhoeffer, T. & Hübener, M. Area-Specific Mapping of Binocular
772 Disparity across Mouse Visual Cortex. *Current Biology* **29**, 2954-2960.e5 (2019).

773 37. Rasmussen, R. N., Matsumoto, A., Arvin, S. & Yonehara, K. Binocular integration of
774 retinal motion information underlies optic flow processing by the cortex. *Current*
775 *Biology* **31**, 1165-1174.e6 (2021).

776 38. Matteucci, G., Bellacosa Marotti, R., Zattera, B. & Zoccolan, D. Truly pattern:
777 Nonlinear integration of motion signals is required to account for the responses
778 of pattern cells in rat visual cortex. *Sci Adv* **9**, (2023).

779 39. Sit, K. K. & Goard, M. J. Distributed and retinotopically asymmetric processing of
780 coherent motion in mouse visual cortex. *Nat Commun* **11**, 3565 (2020).

781 40. Graziano, M. S. A., Cooke, D. F. & Taylor, C. S. R. Coding the Location of the Arm by
782 Sight. *Science (1979)* **290**, 1782–1786 (2000).

783 41. Amedi, A., Malach, R., Hendler, T., Peled, S. & Zohary, E. Visuo-haptic object-
784 related activation in the ventral visual pathway. *Nat Neurosci* **4**, 324–330 (2001).

785 42. James, T. W. & Kim, S. Dorsal and Ventral Cortical Pathways for Visuo-haptic
786 Shape Integration Revealed Using fMRI. in *Multisensory Object Perception in the*
787 *Primate Brain* 231–250 (Springer New York, New York, NY, 2010).
788 doi:10.1007/978-1-4419-5615-6_13.

789 43. Lacey, S. & Sathian, K. Visuo-haptic multisensory object recognition,
790 categorization, and representation. *Front Psychol* **5**, (2014).

44. Maunsell, J. H. R., Sclar, G., Nealey, T. A. & DePriest, D. D. Extraretinal representations in area V4 in the macaque monkey. *Vis Neurosci* **7**, 561–573 (1991).

45. Hagen, M. C. *et al.* Tactile motion activates the human middle temporal/V5 (MT/V5) complex. *European Journal of Neuroscience* **16**, 957–964 (2002).

46. Wang, Q., Sporns, O. & Burkhalter, A. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *Journal of Neuroscience* **32**, 4386–4399 (2012).

47. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72**, 1040–1054 (2011).

48. Tafazoli, S. *et al.* Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *Elife* **6**, (2017).

49. Froudarakis, E. *et al.* Object manifold geometry across the mouse cortical visual hierarchy. doi:10.1101/2020.08.20.258798.

50. Caron-Guyon, J. *et al.* Neuronal encoding of multisensory motion features in the rat associative parietal cortex. *Cerebral Cortex* **30**, 5372–5386 (2020).

51. Ghazanfar, A. A. & Schroeder, C. E. Is neocortex essentially multisensory? *Trends in Cognitive Sciences* vol. 10 278–285 Preprint at https://doi.org/10.1016/j.tics.2006.04.008 (2006).

52. Keller, A. J., Roth, M. M. & Scanziani, M. Feedback generates a second receptive field in neurons of the visual cortex. *Nature* **582**, 545–549 (2020).

53. Stein, B. E. & Stanford, T. R. Multisensory integration: Current issues from the perspective of the single neuron. *Nat Rev Neurosci* **9**, 255–266 (2008).

54. Gemini Team *et al.* Gemini: A Family of Highly Capable Multimodal Models. (2023).

55. Fu, L. *et al.* A Touch, Vision, and Language Dataset for Multimodal Alignment. (2024).

56. Wu, J., Gan, W., Chen, Z., Wan, S. & Yu, P. S. Multimodal Large Language Models: A Survey. (2023).

57. Goh, G. *et al.* Multimodal Neurons in Artificial Neural Networks. *Distill* **6**, (2021).

58. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

59. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* 061507 (2016) doi:10.1101/061507.

60. Wallace, M. T., Meredith, M. A. & Stein, B. E. Multisensory Integration in the Superior Colliculus of the Alert Cat. *J Neurophysiol* **80**, 1006–1010 (1998).

61. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nat Neurosci* **9**, 690–696 (2006).

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

63. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* **8**, 1263–1268 (2005).

64. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap*. (Chapman and Hall/CRC, 1994). doi:10.1201/9780429246593.

839 # Methods

840 Animals

841 All experiments were carried out in accordance with the Institutional Animal Care and
842 Use Committee of the University of Geneva and with permission of the Geneva
843 cantonal authorities (GE258B). C57BL/6J wild-type mice and transgenic mouse lines
844 were housed 2–6 mice per cage under a 12/12-h non-inverted light/dark cycle with ad
845 libitum access to food and water. Transgenic mice used for calcium imaging were
846 obtained as a crossing between Ai148-D mice (Jackson Laboratories, stock number
847 030328) and Rasgrf2-2A-dCre mice (Jackson Laboratories, stock number 022864).
848 Ai148-D mouse line is a Cre-dependent reporter line containing a gene encoding the
849 calcium indicator GCaMP6f at the Igs7 locus. Exposure to Cre recombinase through
850 viral vector injections or crossing with Cre-expressing mice resulted in expression of
851 GCaMP6f. Rasgrf2-2A-dCre mouse line expressed a trimethoprim-inducible Cre
852 recombinase directed by endogenous Rasgrf2 promoter/enhancer elements. When
853 induced, Cre recombinase activity is observed in cortical layers 2/3 and other scattered
854 cells of the cortex, hypothalamus, thalamus, and midbrain. Trimethoprim (TMP) i.p.
855 injections were performed for 5 consecutive days at least two weeks before any
856 surgical intervention (0.25 mg/g of body weight diluted in DMSO and 0.9% NaCl). As
857 a result, TMP injected crossed Ai148D x Rasgrf2-2A-dCre mice expressed GCaMP6f
858 in cortical layer 2/3. For optogenetic experiments, we used Ai32 transgenic mice that
859 express Channelrhodopsin-2 (ChR2) fused to enhanced Yellow Fluorescent Protein
860 (eYFP) in a Cre-dependent manner, enabling fast neuronal activation in vivo through
861 blue light illumination (Jackson Laboratories, stock number 012569). Males and
862 females aged 2–5 months and weighing approximately 20–25 g were used for the
863 experiments.

864 Viral vectors and markers

865 For anterograde labeling, AAV2.CAG.GFP (Addgene, #37825-AAV2, titer: 7×10^{12}
866 vg/mL) and AAV2.CAG.tdTomato (Addgene, #59462-AAV2, titer: 4×10^{12} vg/mL) were
867 used. The volume of viral vectors injected was around 50-75nL in each site at a depth
868 of approximately 400-500 µm from the surface of the cortex. Brains were collected 3
869 weeks after the injections. For retrograde labeling, Cholera Toxin subunit B conjugated
870 to Alexa Fluor 555 (Invitrogen, reference number C22843) or Alexa Fluor 647
871 (Invitrogen, reference number C34778) diluted in PBS were injected (~50nL in each
872 site, 400-500 µm below the surface). Brains were collected 10 days after the injections.
873 For silencing experiments, AAV-DJ.CMV.eGFP-2A-TeNT viral vectors were injected
874 in the region of interest (Wu Tsai Neurosciences Institute, reference number GVVC-
875 AAV-70) at 3 different depths along the cortical column (800 µm, 500 µm and 200 µm
876 respectively). Behavioral protocols started at least one month after the injections were
877 done. For optogenetic experiments, Ai32 mice expressing Cre-dependent ChR2-eYFP
878 were injected with AAV1.CamKII0.4.Cre.SV40 (Addgene, #105558-AAV1,
879 titer: 3.5×10^{12} vg/mL) in three different locations along RL and two different depths
880 (250 µm and 500 µm) to express ChR2-eYFP in most of this area.

881

882 Stereotaxic surgeries

883 Pain management was first performed by administrating the opioid Buprenorphine
884 subcutaneously (0.1 mg/kg) before starting the surgery. Mice were anesthetized inside
885 an induction chamber with 3% isoflurane mixed with oxygen and then fixed on the
886 stereotaxic apparatus (Model 940, Kopf). A custom-made nose-clamp has been
887 adapted to the apparatus to maintain the position of the animal, allowing head rotation.

888 Body temperature was constantly monitored through a thermic probe and adjusted to
889 ~37°C via a heating pad placed below the mouse (DC Temperature Controller, FHC).
890 Breathing rate was regularly monitored by visual inspection. Ophthalmic gel (Vitamin
891 A, Bausch Lomb) was applied on both eyes to ensure protection from light and prevent
892 them from drying out. A local anesthetic was injected under the skin of the head before
893 the surgical incision (mix Lidocaine/Bupivacaine, 6mg/kg and 2.5 mg/kg respectively).
894 During the surgery, mice were anesthetized with a constant isoflurane level lowered
895 around 2%. At the end of the surgery, an anti-inflammatory was also administered
896 subcutaneously (Carprofen, 7.5mg/kg) and animals were warmed with a heating lamp
897 for at least 15 minutes until recovery from anesthesia. A second anti-inflammatory
898 (Ibuprofen, Algifor) was added to drinking water for 3-days post-op and the weight was
899 checked daily to ensure that weight was kept above 15% of the original weight prior to
900 the surgery. All animals were implanted with a cranial implant. After removing the skin
901 and tissues on top of the head, the skull was cleaned, dried and thinned. The mouse
902 head was tilted approximately at a 30° angle, ensuring better access to the left
903 hemisphere. A custom-made metallic implant was placed on the top of the skull using
904 a custom-made holder. It was then fixed with a layer of glue (Loctite 401, Henkel) and
905 additional layers of dental acrylic (Pala, Kulzer) to ensure the implant is securely
906 attached. Dental acrylic was covered with black nail polish to prevent light
907 contamination from visual stimuli during imaging experiments. After at least 3 days of
908 recovery, mice could undergo additional procedures. Animals that underwent imaging
909 sessions were implanted with a glass window composed of a top round cover slip of 7
910 mm diameter and two superimposed 5 mm diameter cover slips (CS-7R and CS-5R,
911 Multi Channel Systems). The three concentric cover slips were glued together with UV
912 glue (Optical Adhesive n°68, Norland). The craniotomy was the size of the smaller
913 window and was drilled above the posterior part of the dorsal cortex. Before removing

914 skullcap, we used a custom-made perfusion chamber with saline solution (NaCl 0.9%)
915 to rinse the craniotomy continuously and reduce bleeding stains. A custom-made
916 holder with air-suction was used to hold and position the window. The cranial window
917 was gently brought down until it was in contact with the brain. The window was then
918 fixed with UV glue, super glue and dental acrylic. For stereotaxic injections, glass
919 pipettes (5-000-2005, Drummond) were pulled (Model P-97, Sutter Instrument Co) and
920 broken to obtain a tip of ~10-15μm inner diameter. Pipettes were further beveled to
921 create a sharp tip to avoid cortical damage during insertion. Injection sites were
922 determined using functional mapping. Injections were done using a single-axis oil
923 hydraulic micromanipulator (R.MO-10, Narishige). The pipette was slowly inserted
924 inside the cortex until reaching the desired depth. Viral vectors or other reagents were
925 injected at a speed of ~2nL/s. When the whole volume was injected, we waited 5
926 minutes with the pipette in the same position before gently removing it.

927

928 Behavioral training

929 A Matlab custom-made graphical user interface (GUI) was developed from the Matlab
930 App Designer to control the behavioral tasks and monitor performance. The GUI
931 allowed real time visualization of the animal's performance and online modification of
932 the parameters (e.g. stimuli parameters, punishment duration, stimuli proportion).
933 Animals underwent water restriction 2 to 4 days before the training started and were
934 handled every day by the experimenter for at least 10 minutes. During the pre-training
935 phase, mice were habituated to head-fixation and placed on the setup. In the first
936 session, only Go stimuli (i.e., stimuli for which licking responses are rewarded) were
937 presented and rewards were delivered automatically regardless of mouse actions. Go
938 trials proportion was progressively reduced as No go trials (i.e., trials featuring the

939 stimulus for which licking responses were punished) were added in the following
940 sessions and mice were required to lick to obtain rewards during Go trials. Mice were
941 trained once a day, every day at the same hour. Eight different behavioral tasks (two
942 tactile tasks, two visual tasks, two auditory tasks and two optogenetic tasks) were used
943 for our experiments (Fig. 1, Extended Data Fig. 1, Fig. 5 and Fig. 6), all following a
944 Go/No go discrimination paradigm. In the tactile tasks, two whiskers ("top" whisker B2
945 and "bottom" whisker C2) were inserted inside glass capillaries each attached to a
946 piezo actuator that could create a small deflection of about 1 mm along the rostro-
947 caudal axis. In one version of the task, B2 whisker stimulations were associated to a
948 reward (Go trials). In the other version, C2 whisker stimulations were the Go trials. In
949 the visual tasks, two drifting squares (a "top" square and a "bottom" square, relative to
950 the midline in the mouse visual field) were presented on the screen. In one version of
951 the task, top square stimulations were the Go trials while in the other version of the
952 task, bottom squares stimulations were the Go trials. The auditory task followed the
953 same structure as the other tasks but using two short pure tones of 6 kHz and 12 kHz
954 as Go and No go stimuli, respectively. These tones were delivered from a speaker
955 located next to the mouse on the same side as the visual and tactile stimuli. The
956 optogenetic task was performed by directly stimulating subregions of RL
957 corresponding to top or bottom multimodal representation using shaped blue light
958 patterns. In one version of the task, stimulating the top region was associated with a
959 reward, while in the other version, it was the bottom region. The total trial duration of
960 a single trial was 4 s: after a 2 s quiet window (during which licking resulted in trial
961 abortion) the stimulus was presented, and the mouse was allowed to lick during a 2 s
962 response window. During Go trials, the mouse could obtain a water reward upon
963 licking the spout following the stimulus (Hit trials). Failure to lick would result in a Miss
964 trial. During No go trials, the mouse had to refrain from licking (Correct rejection trials,

965 CR) or it was punished with a time-out of 10 s (False Alarm trials, FA). Some trials
966 were presented without any stimulus (catch trials). If the mouse licked during catch
967 trials, no time-out was applied. If mice were too compulsive (i.e. licking during the quiet
968 window was too frequent), a 10 s time-out early lick punishment could also be applied.
969 The proportion of each trial type was the following: Go trials = 30%, No go trials = 50%,
970 Catch trials = 20%. All tactile stimuli were generated through Matlab data acquisition
971 toolbox controlling a piezo actuator (Bimorph bendor piezo actuator PB4NB2S,
972 Thorlabs) through a National Instrument card. All visual stimuli were generated using
973 Matlab and PsychToolBox. Stimuli were presented on a gray background through a
974 LCD monitor (20x15cm, pixels, 60Hz refresh rate, Pi-shop) positioned 10 cm from the
975 eye, with a 30° angle to the right of the midline. The screen was also tilted with a 30°
976 angle along the horizontal plane to match the mouse head angle with the intent of
977 roughly aligning the bottom and top parts of the screen to the resting position of the
978 C2 and B2 whiskers in the mouse visual field. Stimulations during behavior consisted
979 of black squares moving through the screen in the rostro-caudal direction (bar width
980 was 15°, stimulus duration was 175ms, speed was 500°/sec), on a gray background.

981

982 Wide-field microscopy

983 We used a custom-made wide-field epifluorescence microscope setup[39] including a
984 sCMOS camera (ORCA-Flash4.0 V3, Hamamatsu). Magnification was achieved
985 through a 0.63X C-mount camera adapter for Olympus Microscopes. The field-of-view
986 size was 5.6 mm x 5.6 mm. The camera and adapters were mounted on a base
987 allowing vertical movement with manual focus. LED white illumination (740mW,
988 1225mA, Thorlabs) could be controlled via a T-Cube LED driver (LEDD1B, Thorlabs).
989 Filter cubes could be changed for different types of imaging. For imaging GCaMP6f,

990 GFP excitation, emission and dichroic filters were used. For intrinsic optical signals
991 imaging, Cy3/5 excitation, emission and dichroic filters were used. An objective (MVX
992 Plan Apochromat with 2x, Olympus) was attached to the microscope base. The
993 somatotopic mapping protocol consisted of repetitive rostro-caudal pulsatile
994 deflections (~1 mm amplitude) of either B2 or C2 whiskers for 50-80 trials each
995 followed by a quiet window. The retinotopic mapping protocol consisted of drifting bars.
996 A contrast reversing checkerboard was presented within the bar to better drive neural
997 activity (0.04 cycles/° of spatial frequency and 2 Hz of temporal frequency). In each
998 trial the bar was swept in the four cardinal directions: left to right, right to left, bottom
999 to top, and top to bottom. Single trials were repeated 20 to 40 times. For anatomical
1000 experiments, only C57BL/6J mice were used. To image the intrinsic optical signal, we
1001 used longer tactile stimulations and slower visual stimulations. Mice were fixed on the
1002 platform and anesthetized during the procedure with isoflurane level lowered than 1%.
1003 Body temperature was monitored with a probe and adjusted to 37°C using a heating
1004 pad (DC Temperature Controller, FHC). For all the other experiments, we used Ai148-
1005 D x Rasgrf2-2A-dCre mice and calcium imaging. The visuo-tactile sparse noise
1006 protocol consisted in combinations of visuo-tactile stimuli: three whisker conditions (C2
1007 whisker, B2 whisker and no whisker), eight vertical positions of the moving square.
1008 The ninth position corresponded to a no visual stimulus condition (i.e. blank screen).
1009 Visual and tactile stimulus onsets were either synchronous or delayed (0.15 s delay,
1010 visual leading tactile stimulus). In total, 54 different combinations were presented in
1011 pseudo-random order with a 1 s inter-stimulus interval. The full sequence was
1012 repeated 60 times. Total sparse noise protocol duration was approximately 1 hour. At
1013 the beginning of each recording, we took a picture of the cranial window with blood
1014 vessels pattern on focus as reference image. The focus was then set at ~300μm below
1015 the surface to maximize signal collection. Light was adjusted to prevent saturation.

1016 Before each imaging session, the window was cleaned with 70% ethanol and eyes
1017 hydrated with mineral oil.

1018

1019 Two-photon microscopy

1020 The two-photon microscope was custom-made (INSS Company). It consisted of a
1021 femtosecond laser with wavelength range 690-1040nm (Tunable Ti:Sapphire with
1022 dispersion compensation MaiTai DeepSee, Spectra Physics) which beam was
1023 directed with Resonant/Galvo scan mirrors and the emitted signals were detected by
1024 2 GaAsP amplified PMTs (PMT2101/M, Thorlabs). Imaging was performed through
1025 ScanImage (Vidrio Technologies). Images were acquired at approximately 30 frames
1026 per second. Two-photon calcium imaging during the visuo-tactile sparse noise
1027 protocol described in the wide-field section was performed on Ai148-D x Rasgrf2-2A-
1028 dCre mice. After cleaning the cranial window with 70% ethanol, a hydrophobic
1029 chamber was made between the head plate and the imaging platform using liquid
1030 plastic (Smooth-Cast 325, Smooth-On) and the objective was immersed in distilled
1031 water. We ensured no polluting light could reach the objective by covering it with a
1032 dark protective sleeve and by turning off the light in the room. At the beginning of each
1033 recording, an anatomical picture of the field-of-view was taken using a CCD camera.
1034 When switching the microscope in two-photon mode, we took an image of the surface
1035 blood vessels at magnification x1 and x1.5 for further realignment. For each location,
1036 we typically imaged at three different depths between 100 µm and 300 µm below the
1037 surface. Each mouse underwent no more than two imaging protocols per day.

1038

1039

1040 Histology

1041 Mice were anesthetized with 3% isoflurane and euthanized with i.p. injection of
1042 Pentobarbital (Eskornarkon, 150 mg/kg) and Buprenorphine (0.1 mg/kg). They were
1043 then transcardiacally perfused using a peristaltic pump (ISM829, Cole-Parmer) with
1044 0.01 M PBS, pH 7.4 for 2 min, and then 4% paraformaldehyde (PFA) in phosphate
1045 buffer (PB; pH 7.4) for 3 min. The brains were post-fixed at 4°C in PFA for 48 h and
1046 then transferred into PBS. For anatomical experiments of Fig. 3, the brains were post-
1047 fixed at 4°C in PFA for 2 h and washed 3 times for 15 min in PBS. The left cerebral
1048 hemisphere was separated from the right hemisphere and subcortical parts were
1049 removed with a spatula. The left hemisphere was flattened between glass slides and
1050 kept flattened at 4°C in PFA for 12 h. After PBS washing, flattened hemispheres were
1051 embedded in agar gel 4% before cutting slices of 60µm thickness with a vibratome
1052 (VT 1000 S, Leica).

1053

1054 Immunohistochemistry

1055 Brain sections were incubated with slight agitation (40 rotations/min) for 2 h at room
1056 temperature in a saturation/permeabilization solution containing a mix of 5% BSA and
1057 NGS, 0.3% triton X-100 and PBS. Brain sections were then incubated with slight
1058 agitation (40 rotations/min) overnight at 4°C with a rat anti-muscarinic acetylcholine
1059 receptor M2 (M2 AChR) primary antibody (1:500, Sigma-Aldrich, reference number
1060 MAB367) in the same blocking solution. Sections were washed 3 times for 15 min in
1061 PBS before a 2 h incubation with either a donkey anti-rat secondary antibody
1062 conjugated with Alexa Fluor 488 (dilution 1:500, Invitrogen, #A21208) or with a goat
1063 anti-rat secondary antibody conjugated with Cy5 (dilution 1:500, Invitrogen, #A10525)

1064 in blocking solution at room temperature without agitation. Hoechst solution (dilution
1065 1:1000, Invitrogen, # 33342) was used for fluorescent nuclear counterstaining. Slices
1066 were incubated for 10 min in that solution with slight agitation then washed 2 times for
1067 5 min in PBS.

1068

1069 Histological imaging

1070 Slices were mounted on Superfrost microscope slides (Epredia) with mounting
1071 medium (Fluoromount) and covered with 24x50 mm coverslips (Menzel-Gläser). All
1072 photomicrographs were taken using a Zeiss Axio Scan.Z1 or a Leica Stellaris 5
1073 microscope at the bioimaging platform of the University of Geneva.

1074

1075 Wide-field calcium imaging analysis: Retinotopic and whisker mapping protocols

1076 Responses to drifting checkerboard stimuli were averaged across trials for each
1077 condition and converted in df/f using a pre-stimulus time window as baseline. Azimuth
1078 and elevation preference maps were then computed[24].To segment visual areas,
1079 azimuth and elevation maps were combined to generate a visual field sign map. The
1080 sign map was computed as the sine of the difference between the vertical and
1081 horizontal retinotopic gradients for each pixel. Somatotopic maps were obtained by
1082 averaging the response df/f across B2 or C2 stimulation trials over a short time window
1083 of 200 ms.

1084

1085

1086 Wide-field calcium imaging analysis: Sparse noise protocol

1087 After an initial down sampling to a resolution of 100 x 100 pixels by bicubic
1088 interpolation, 50 Hz framerate videos acquired during the sparse noise protocol
1089 underwent pixelwise notch filtering (f_0=12 Hz, f_w=6 Hz) and discrete wavelet transform
1090 (DWT) detrending (setting lowest frequency approximation coefficients to zero in a 6-
1091 level decomposition) to remove artifacts and low-frequency drifts. After these initial
1092 preprocessing steps, videos were z-scored using as reference distribution the
1093 ensemble of all pixel values corresponding to blank stimuli (i.e. trials with no visual
1094 and no tactile stimulation). Next, trials with a high baseline activity (>=0.5 z-score on
1095 average, corresponding to a strong positive fluctuation of spontaneous activity
1096 preceding the stimulus) were discarded and z-score and baseline subtraction were
1097 computed again including only remaining trials. This was meant to better separate
1098 local responses to sensory stimulations from ongoing spontaneous activity. Starting
1099 from the z-scored data obtained in this way, response surprises (i.e. $-\log10(p_{t\text{-test}})$)
1100 across trials were computed for each time point and each stimulus to define spatial
1101 responsivity masks. Average visual or tactile top or bottom response movies were
1102 obtained by averaging the median responses across trials to the relevant stimuli. For
1103 visual "top" all stimulation conditions (either unimodal or multimodal) in which the
1104 visual stimulus was present in grid positions 7 or 8 were used whereas for visual
1105 "bottom" all visual conditions including positions 3 or 4 were used. Retinotopic and
1106 somatotopic maps of vertical space (Fig. 2b,c and Extended Data Fig. 5g) were
1107 obtained by taking the difference between average visual or tactile top and bottom
1108 response maps. Modality preference maps (Fig. 2d and Extended Data Fig. 5g) were
1109 obtained by computing pixel-by-pixel the difference between the maximum response
1110 for visual and tactile stimuli. Multisensory enhancement maps (Fig. 2f and Extended

1111 Data Fig. 5g) were obtained by computing pixel-by-pixel the difference between the
1112 maximum unisensory response between visual and tactile stimuli (predicted response)
1113 and the multisensory response (measured responses). For all the difference maps
1114 mentioned above the input maps were independently normalized to the [0,1] interval
1115 before taking the difference (ensuring their range to be bound in the [-1,1] interval).
1116 Retinotopy-somatotopy spatial coherence maps (Fig. 2e and Extended Data Fig. 5g)
1117 were obtained by convolutionally (stride = 1 patch size = 8 pixels) computing the
1118 correlation-based similarity of matching patches of the two different maps. The result
1119 was then scaled by the average peak-to-peak range in that patch (to diminish
1120 contribution from pixel with very low response magnitude). Comparison of grand
1121 average maps at single pixel level (Fig. 2g,h) was performed on responsive pixels only
1122 (exceeding a surprise threshold corresponding to $p_{t\text{-test}} <= 0.1$ for either visual or tactile
1123 responses). We compared pixels with low and high modality preference (i.e. below or
1124 above 85% quantile of the absolute value distribution within the responsive region) or
1125 with low and high coherence (i.e. below or above 15% quantile of the absolute value
1126 distribution within the responsive region). Grand average maps displayed in Fig. 2b,c,f
1127 are obtained by averaging frames of the corresponding grand-average movie
1128 (obtained by realigning single mouse movies to the common reference atlas and then
1129 averaging) over different time windows: frames 18 to 20 for the somatotopy (i.e. 0.16
1130 s to 0.20 s after stimulus onset); frames 25 to 27 for the retinotopy (i.e. 0.30 s to 0.34
1131 s after stimulus onset); and frames 28 to 30 for multisensory modulation (i.e. 0.36 s to
1132 0.4 s after stimulus onset). These windows were chosen to be roughly centered on the
1133 peak of these signals. Grand average modality preference and spatial coherence
1134 maps displayed in Fig. 2d,e were computed on maps obtained above (average z-score
1135 maps or position selectivity maps, respectively). Signal time courses displayed in
1136 Extended Data Fig. 6 were obtained by integrating grand-average movies across time-

varying regions of interest (ROIs). These ROIs were specifically defined to track the activity 'bump' evoked by stimulation in each area. To prevent bias in estimating the response —particularly in areas with higher cortical magnification factors, which could lower the average by including many non-responsive pixels— these masks were created by intersecting a static mask, which selects each area based on the fitted reference atlas, with a dynamic responsivity mask. This dynamic mask selects, for each movie frame, only those pixels that across mice show an average maximum response surprise, exceeding a threshold (corresponding to a p-value ≤ 0.01 across conditions). To measure angular mismatch between retinotopic and somatotopic maps we computed vertical retinotopy gradient vectors from the grand-average version of these maps (using Matlab “gradient” function with a scale of 4 pixels followed by a Gaussian smoothing of gradient components). Then a region of interest overlapping with RL was defined and we computed the average angular difference between the two gradient fields.

<u>Atlas fitting and registration</u>

The reference Allen Mouse Brain atlas[58] was first tilted sideways at 30° and then projected to match the skull tilt in our experiments and therefore the layout of areas in the dorsal cortex. The resulting atlas was manually fitted to the wide-field imaging field-of-view for each mouse by visually aligning the atlas boundary lines to reproducible landmarks from functional maps. Both the maps obtained from the sparse noise protocol (i.e. vertical retinotopy and somatotopy difference maps) and the maps obtained from the whisker and retinotopic mapping protocols were used to register the atlas. Landmarks used include the outlines of the sign map regions, the position of C2 and B2 whisker activity bumps in S1 and S2, the reversal of vertical retinotopy at the

1162 boundary of each visual area (see Extended Data Fig. 5 for example). To relate
1163 microscale and macroscale functional properties measured respectively by two-
1164 photon and wide-field imaging experiments, we also manually reconstructed the
1165 position of each two-photon field-of-view (FOV) in the frame of reference with the atlas
1166 fitted to the wide-field maps. To do so, we aligned the blood vessels pattern visible in
1167 each two-photon FOV with the one visible in the wide-field cortical image (as shown
1168 in Extended Data Fig. 8f). To compare data across mice and compute grand averages,
1169 we developed a Matlab pipeline to robustly realign atlases to one another. This
1170 pipeline was based on iterative application of the image registration algorithm
1171 implemented by Matlab function "imregtform" (considering "rigid" transformations in
1172 "monomodal" mode) to stack images depicting atlas boundaries. This pipeline enabled
1173 us to obtain the rotation and shift required to register all maps and neuron positions to
1174 the frame of reference of a common atlas. For the anatomical tracing experiments (Fig.
1175 3b-d), a similar processing was performed with ImageJ through registration of atlases
1176 fitted using the M2 AChR staining. Brain from all mice with similar injections could then
1177 be averaged together. To compare the somatotopy and retinotopy of axonal
1178 projections or projecting neurons with the functional topographic maps obtained in
1179 wide-field we brought them in a common reference frame by registering a crop of the
1180 M2 AChR-fitted atlas to the common wide-field/two-photon atlas (by finding the
1181 registering affine transformation with Matlab functions "cpselect" and "fitgeotrans").

1182

1183 <u>Two-photon calcium imaging analysis: Pre-processing and functional cell-types</u>

1184 To extract time-varying somatic GCaMP6f calcium signals, we used the Suite2p
1185 toolbox[59]. Neuropil contamination was corrected by subtracting the fluorescent signal
1186 from a surrounding ring $F_{Surround}(t)$ from somatic fluorescence: $F(t) = F_{Soma}(t)$ -

1187 $\alpha * F_{Surround}(t)$ with $\alpha=0.7$. Neuropil-corrected fluorescence signals F(t) were then
1188 converted in z-score by subtracting from each trace the mean value and dividing by
1189 the standard deviation of F(t) over the last 0.2 s of the baseline window preceding the
1190 stimulus onset (pooling across all trials). Using these z-scored fluorescence traces we
1191 computed response surprises (i.e. $-\log10(p_{z\text{-}test})$) across trials for each stimulus and
1192 for each time sample. Similarly, we computed the coefficient of variation (CV) across
1193 trials quantifying response variability. A neuron was considered reliably responsive if
1194 its response exceeded a stringent surprise threshold $Surprise_{th} = 8$ (i.e. $p_{z\text{-}test} <= 10^{-8}$)
1195 while remaining below a coefficient of variation threshold value $CV_{th} = 4$ for at least 4
1196 consecutive time bins in any stimulus condition. This responsivity criterion was
1197 requested for each neuron to be included in subsequent analyses. For each neuron
1198 included, we used a more inclusive criteria to characterize the extent of their tuning
1199 properties. Indeed, neurons exceeding a less stringent threshold $Surprise_{th} = 2$ (i.e. $p_{t\text{-}}$
1200 $_{test} <= 0.01$) while remaining below the same coefficient of variation threshold at any
1201 time bin of a visual, tactile or visuo-tactile condition were considered “responsive” to
1202 these conditions. The responsivity pattern of each neuron was used to classify them
1203 in one of four functional classes: 1) neurons responsive in visual but not tactile
1204 conditions were termed “visual neuron”; 2) neurons responsive in tactile but not visual
1205 conditions were termed “tactile neuron”; 3) neurons responsive in both tactile and
1206 visual conditions were termed “visuo-tactile neuron”; 4) neurons responsive only in
1207 visuo-tactile conditions (both visual and tactile stimuli presented together) were termed
1208 “gated neuron”. Only visuo-tactile neurons displaying significant responses in both
1209 modalities were used for the analyses displayed in Fig. 4g-j. Measured multisensory
1210 responses were compared to a “max model” of multisensory interaction[60] (as in Fig.
1211 4b and Extended Data Fig. 10b). Predicted visuo-tactile z-scored traces for each
1212 multisensory stimulus condition were computed trial-wise by taking the max (for each

1213 time bin) between the response to the same visual and tactile stimulus presented
1214 alone (i.e. the corresponding unimodal conditions) and subsequently averaging across
1215 trials. Average responses (predicted or measured) were computed over a response
1216 window spanning from 0.2 s to ~0.733 s following stimulus onset for each stimulus
1217 and each responsive neuron.

1218

1219 Two-photon calcium imaging analysis: Computation of indices

1220 We computed the "multisensory modulation index" (MI) as the normalized difference
1221 between predicted and measured visuo-tactile responses at every time bin for every
1222 stimulus condition. To summarize the intensity of multisensory modulation of each
1223 neuron across stimulus conditions over time we computed the min-max range of the
1224 integral of MI values over the same time window mentioned above (i.e. from 0.2 s to
1225 0.733 s) across all valid conditions. Valid stimulus conditions comprised significant
1226 responses as well as all multisensory conditions containing at least one unisensory
1227 stimulus to which the neuron is responsive and displaying significant multisensory
1228 modulations (i.e. with $p<0.05$ of a bootstrap-t test for MI value for at least one time bin,
1229 see "quantification and statistical analysis" section below). Preferred vertical position
1230 in the tactile or visual space of each neuron was quantified as the center-of-mass of
1231 these average responses along the corresponding dimension of the stimulus grid on
1232 visuo-tactile (measured or predicted) conditions. This center-of-mass was converted
1233 into a "position preference index" ranging from +1 for neurons tuned to the top
1234 visual/B2 whisker stimulus to -1 for neurons tuned to the bottom visual/C2 whisker
1235 stimulus (as in Fig. 4c,d,g and Extended Data Fig. 9a-b). The sharpness of positional
1236 tuning of neurons was computed as the inverse of the best sigma parameter of a
1237 Gaussian fit of the average response to each visual position or as the absolute value

1238 of the tactile position selectivity index described above. This selectivity index was
1239 computed on the visuo-tactile condition of max response (both in the measured or
1240 predicted case). These indices ranged from 0 for neurons responding equally to all
1241 positions (i.e. completely positionally untuned) to 1 for neurons responding to only one
1242 position (i.e. maximally positionally tuned). The values of positional preference and
1243 selectivity indices shown in Fig. 4g,h,i,j are averaged over both delay conditions (i.e.
1244 with or without delay between tactile and visual stimuli, see Extended Data Fig. 6).
1245 Neurons were labeled as ventral or dorsal by running a k-means clustering algorithm
1246 (with $n_{centroids}$=2) on the spatial distribution of visuo-tactile neurons over the surface of
1247 the dorsal cortex (Fig. 4e).

1248

1249 Two-photon calcium imaging analysis: Correspondence with wide-field

1250 To quantify the correspondence between single-neuron properties distributed across
1251 the cortex and the wide-field maps, we created smoothed spatial maps from 2D
1252 histograms that depict the dominant response properties at each cortical location. We
1253 generated retinotopic and somatotopic maps from two-photon data (Fig. 4c,d and
1254 Extended Data Fig. 9a,b) by smoothing a 2D normalized histogram of neurons with
1255 positional preferences (top-preference index≥0.05, bottom-preference index≤0.05)
1256 using a Gaussian kernel (σ~3% of total wide-field FOV) and then calculated the
1257 difference between these maps. For modality preference maps reconstructed using
1258 two-photon data (Extended Data Fig. 9c), we applied a Gaussian smoothing (σ~3% of
1259 FOV) to 2D normalized histograms that describe the distribution of unimodal visual
1260 and tactile neurons, calculating the normalized difference between them. To map
1261 multisensory modulation (Extended Data Fig. 9d), we applied Gaussian smoothing
1262 (σ~3% of FOV) to a MI-range-weighted histogram showing the spatial distribution of

1263 multisensory modulations, then normalized this against the unweighted distribution
1264 that includes unresponsive neurons. We included unresponsive neurons to properly
1265 capture the relative density of multisensory modulation in the local population. To
1266 quantitatively assess the match between two-photon reconstructed maps and the
1267 wide-field maps, we computed a Pearson correlation coefficient restricted to the pixels
1268 with good two-photon coverage (thresholding the map shown in Extended Data Fig.
1269 8g at a threshold of 2,000 neurons/mm2).

1270

1271 Two-photon calcium imaging analysis: Decoding analysis

1272 To test whether the neuronal representation observed across the dorsal cortex of
1273 naïve mice could support cross-modal generalization observed in our behavioral
1274 experiments, we adopted a decoding approach. We trained a Bayesian classifier to
1275 infer stimulus category (top or bottom) from single-neuron responses in the tactile
1276 modality and then tested the accuracy on the visual modality. This decoder was
1277 inspired by original works introducing neuronal population probabilistic codes[61,62] and
1278 is similar to the approach used in previous work[20]. We built feature vectors
1279 representing the response of each neuron to each stimulus category (i.e. “top” and
1280 “bottom”) in each modality. The feature vectors were defined, for each neuron, as the
1281 z-score response integrated over a 0.2 s - 0.733 s time window in every trial of the
1282 most responsive stimulus position pertaining to each category (visual top = positions
1283 5 to 8, visual bottom = positions 1 to 4, tactile top = B2, tactile bottom = C2). The
1284 response feature vectors were used in a trial-wise fashion as predictors to infer the
1285 “top” or “bottom” labels corresponding to the location of the stimulus delivered to the
1286 animal. We incorporated 4-fold cross-validation to mitigate overfitting and assess the
1287 classifier's accuracy outside the training dataset. Trials were partitioned into four

batches, three of them for training and one for testing. The decoding was assessed both at the single neuron level (n_{pool} = 1 and averaged over all neurons of a given class to get the values displayed in Extended Data Fig. 10d) as well as on pools of neurons of increasing size (n_{pool} = [1,50] to get the curves displayed in Extended Data Fig. 10c). During population decoding, we repeated the procedure 40 times, each time randomly resampling the neurons in the decoding pool without replacement to ensure that the results were not dependent on a specific selection and partitioning of neurons. Decoding accuracy (defined as the fraction of trials with correct predictions) was averaged across resampled data and cross-validation splits. Prior to classification, each neuron's response features within the population vector were individually normalized to a 0-1 range across trials. Additionally, trials were randomly reordered, maintaining their corresponding correct labels, to eliminate noise-related correlations. In this probabilistic framework the ‘‘training” consists simply in computing the average tuning curves (i.e. stimulus-triggered average) using all trials included in the training set and taking the logarithm of this vector. The subsequent ‘‘testing’’ of the decoder requires to generate a prediction. Predictions were made by calculating each test trial's feature vector log-likelihood for the 'top' or 'bottom' stimulus, then assigning the label with the highest log-likelihood as the predicted outcome. The log-likelihood can be obtained by taking a matrix product between the matrix storing the log of tuning curves computed at the ‘‘training’’ stage and the current population feature vector (after subtracting a corrective term proportional to the sum of all tuning curves to account for any representational bias in the decoded population). Tactile decoding performances represented in Extended Data Fig. 10c,d corresponded to the average cross-validated test accuracies on the same modality on which the Bayesian classifier was trained (i.e. tactile).

1313

1314 Behavioral analysis

1315 To quantify mouse behavior over time within single sessions we isolated different
1316 stimulation conditions (i.e. trials with top stimuli, trials with bottom stimuli and Catch
1317 trials) to compute a smoothed conditional lick probability over a sliding window of 60
1318 trials. We calculated the conditional probabilities of licking by dividing the number of
1319 trials with a mouse lick response to a specific stimulus by the total presentations of
1320 that stimulus within the moving time window. These curves are the ones displayed
1321 throughout the paper (Fig. 1,5,6 and Extended Data Fig. 1,2,3,11,12) and labelled
1322 "Top", "Bottom" and "Spont". From Go and No go lick probabilities (defined by the task
1323 rule to coincide with top or bottom), we also computed a "Discrimination performance"
1324 metric as the percentage of correct trials for discrimination. Specifically, we defined
1325 behavioral performance as the average between the rate of correct choices upon
1326 delivery of the Go stimulus (i.e. Hit rate = conditional lick probability for Go trials) and
1327 the rate of correct rejects upon delivery of the No go stimulus (i.e. Correct Reject rate =
1328 1 — False Alarm rate where False Alarm rate = conditional lick probability for No go
1329 trials). This way of computing the performance makes it insensitive to the proportion
1330 of Go and No go trials. Following a similar logic, we defined a "Detection performance"
1331 as the maximum between the Go detection rate (i.e. average Hit rate and Catch
1332 Correct Reject rate) and the No go detection rate (i.e. average False Alarm rate and
1333 Catch Correct Reject rate). Behavioral states were classified as previously described[20].
1334 Conditional lick probabilities and performance per session were computed only during
1335 engaged trials of each session (i.e. discarding the part of the session during which
1336 conditional lick probabilities dropped due to task disengagement). Switch-aligned
1337 performance trajectories shown throughout the paper (Fig. 1,5,6 and Extended Data

Fig. 1,3,12) were obtained realigning the curve of average session performance and conditional lick probabilities around the session of task switch. Accompanying bar plots display the difference of discrimination and detection performance between the last session before the switch and the first session after the switch. In Extended Data Fig. 2, we plotted the average conditional lick probabilities and the performance time course across mice that underwent whisker removal control sessions. These average curves were obtained by dividing each session in three chunks: “before”, “during” and “after” whisker removal and averaging them separately after a stretching interpolation to match their duration. Accompanying bar plots display the average discrimination and detection performance in these three chunks (excluding boundary timepoints to avoid contamination due to the temporal smoothing induced by the sliding window we used to compute the traces). Learning rates following the switch (Fig. 1h and Fig. 5l) were obtained measuring the slope of the best fitted line for 3 consecutive sessions after switch for all mice in each group. We excluded mice that did not show a drop of performance to chance level after the switch as they successfully generalized from one modality to the other. The relationship between the TeNT-P2A-GFP expression overlap with each cortical area and the cross-modal generalization performance impairment in rule-preserving conditions was assessed using Pearson coefficient of correlation. Specifically, for each mouse we calculated the Pearson coefficient of correlation between two vectors: 1) the vector representing the fractional overlap of TeNT expression binary masks (displayed in Extended Data Fig. 11) and each atlas-defined cortical area, and 2) the vector representing the change in performance following modality switch (performance averaged over 3 consecutive sessions after switch). This analysis was performed for each cortical area (examples for areas RL and AL are shown in Extended Data Fig. 11f) and displayed as a correlation map in Fig. 5e. For the reverse correlation approach, we computed the difference between

1364 the average of all TeNT-injected mice GFP fluorescence binary masks and the one of
1365 mice displaying impairment of cross-modal generalization ability (defined as mice
1366 performing below 57.5% of correct trials on average in the 3 sessions following a rule-
1367 preserving modality switch). This difference took null or negative values in areas where
1368 TeNT expression did not affect the behavior and positive values in regions producing
1369 behavioral impairment (i.e. regions where TeNT expression is enriched in the
1370 impairment-triggered ensemble compared to the ensemble including all TeNT-
1371 expressing mice).

1372

1373 Quantification of TeNT expression and ventral/dorsal grouping

1374 Wide-field fluorescence images of dorsal cortex displaying the pattern of expression
1375 of TeNT-P2A-GFP in each injected mouse were first pre-processed to remove the
1376 blood vessels from the surface. This was necessary to obtain a smooth, non-occluded
1377 estimate of the spatial profile of GFP expression across the surface of the cortex. To
1378 do so we first segmented the blood vessels with adaptive thresholding (using Matlab
1379 function “adaptthresh”) and then inpaint them by local interpolation (using Matlab
1380 function “regionfill”) (Fig. 5g,i and Extended Data Fig. 11a-d and 12b). Next, we
1381 computed a median-subtracted map of df/f calculated with respect to a manually
1382 annotated region of interest on the border of each window (selected to correspond to
1383 a fluorescence profile stable around baseline level, usually at the farthest end of the
1384 window with respect to the injection site). To compare images across different mice
1385 and account for variations in imaging conditions, we normalized saturation by clipping
1386 the values in each map to the minimum peak value observed across all mice. After
1387 normalizing these resaturated df/f maps to 1 we computed their 0.90 contour (i.e. the
1388 expression “patch”) to define the binary TeNT-P2A-GFP expression mask to be used

1389 in subsequent analyses to evaluate the spatial extent of toxin expression. The overlay
1390 of these masks for all mice is displayed in Extended Data Fig. 11e. Next, TeNT-
1391 expressing mice undergoing rule-preserving modality switch were subdivided in
1392 "dorsal" and "ventral" subgroups by computing a d-prime measure (distance
1393 normalized by patch size) with respect to the two multimodal visuo-tactile neuron
1394 clusters found by k-means in the two-photon analysis of Fig. 4e. Neurons outside the
1395 region of low d-prime with patch centroid falling respectively within AL/LI or A/RL
1396 region were categorized as ventral and dorsal, respectively (Fig. 5g,i showed the
1397 overlay of the binary masks together with expression centers as dots over the common
1398 atlas). The same procedure was followed to validate the targeting of TeNT injections
1399 to the dorsal region in mice undergoing rule-reversing modality switch (Extended Data
1400 Fig. 12b).

1401

1402 Anatomical tracing analysis

1403 For each brain, M2 AChR immunostaining and tracer fluorescence images of each
1404 slice were realigned by matching the radial blood vessel pattern. To do so, we applied
1405 the "Normalize local contrast" function in ImageJ and then inverted the image to crop
1406 low intensity pixels and isolate blood vessel holes. We then iteratively registered these
1407 images across the stack in ImageJ using the "MultiStackReg" function with affine
1408 transformation. The transformation could then be applied to the original stack. Slices
1409 with good fluorescence signals were then averaged together. For anterograde tracing
1410 data, we first applied an "Unsharp mask" function (radius: 20, mask weight: 0.8) and
1411 pixel values were log-transformed to enhance axonal projections with respect to the
1412 strong signal at the injection site where cell bodies are labelled. For retrograde tracing
1413 data, a top hat filter (radius: 12) was applied to highlight the signal carried by CTB-

1414 positive cells followed by the “Enhance contrast” function (Saturated pixels: 0.55%).
1415 Realigned average slices for each mouse were then manually fitted to a projection of
1416 the Allen mouse brain atlas using M2 AChR landmarks[58]. Individual mouse atlases
1417 were then registered to common reference atlas. Centroid positions for injection
1418 locations were determined by calculating the center of mass of the fluorescence
1419 signals within manually outlined regions surrounding the injection sites. For
1420 anterograde tracing data, fluorescence images corresponding to different injections
1421 were normalized to the unit range per channel across brains and median-equalized
1422 across channels within brains. For retrograde tracing data, fluorescence images
1423 corresponding to different injections were rescaled to have the same peak value
1424 across channels and brains. Center-of-mass coordinates for each injection in each
1425 brain were projected on a line corresponding to the vertical meridian in the somatotopic
1426 or retinotopic map at the targeted site. This yielded, for each injection, a vertical
1427 coordinate as the center location. Lastly, we obtained an average projection map by
1428 weighting the coordinate of each injection center with its normalized pixel intensity at
1429 every pixel across the stack of fluorescence images (Fig. 3b,c,d and Extended Data
1430 Fig. 7g). The maps obtained in this way represent for every pixel the preferred
1431 projection origin (anterograde) or projection target (retrograde) location along the
1432 vertical line, allowing for an understanding of the topographic mapping of inputs and
1433 outputs associated with various regions. We also established a statistical metric to
1434 assess the significance of projections at each location. This was accomplished by
1435 comparing the median difference between two distributions: the local distribution of
1436 normalized pixel intensities at each location and the distribution of normalized pixel
1437 intensities from a specific frontal-lateral region of the slice. The latter region was
1438 selected based on its lack of fluorescence signal, as determined by an initial visual
1439 examination of the raw images. The center of the color scale used to plot the projection

1440 maps corresponds to the median projection coordinate of the injection ensemble.
1441 Average projection maps were compared to wide-field retinotopic and somatotopic
1442 maps. This was achieved by aligning a selected portion (matching the cranial window
1443 field-of-view) of the projection maps with the shared wide-field/two-photon atlas.
1444 Subsequently, we calculated the Pearson coefficient of correlation between the two
1445 images as a measure of similarity. Pixels exhibiting non-significant projection signal
1446 were discarded for this analysis.

1447

1448 Optogenetic substitution experiments

1449 To test if the dorsal cortical area RL is sufficient for cross-modal generalization, we
1450 used area-specific optogenetic activation as a substitute for visually evoked responses.
1451 We expressed Channelrhodopsin-2 in the cortex by injecting
1452 AAV1.CamKII0.4.Cre.SV40 into the whole RL area of Ai32 mice. This approach
1453 induced a widespread expression of Channelrhodopsin-2 (ChR2) fused with
1454 Enhanced Yellow Fluorescent Protein (eYFP). ChR2, a light-gated cation-selective
1455 channel, generates depolarizing photocurrents in response to blue illumination,
1456 thereby allowing for the excitation of cortical neurons[63]. To illuminate the cortical
1457 surface of head-fixed mice during behavioral tasks, we built a behavioral rig equipped
1458 with a Digital Light Processing (DLP) module (STAR 3.0 EVM Monochrome LED
1459 Projection Modules, DLP6500, ViALUX GmbH). This module features a 1024 × 768
1460 Texas Instruments micro-mirror chip and a 460 nm blue LED, integrated with an
1461 inverted tandem-lens configuration macroscope. The optical tower assembly
1462 comprised two lenses configured in an L-shaped pathway, separated by a dichroic
1463 mirror. Light from the macroscope enters through the projection lens (L1; 150 mm smc
1464 Pentax-A 645 lens) at the rear, reaching a dichroic mirror (640 nm cutoff wavelength;

1465 60 × 60 mm; Semrock) positioned at a 45° angle. This mirror reflects the light towards
1466 the front end of the objective lens (L2; 50 mm Nikon NIKKOR lens), which projects the
1467 image onto the mouse's cortex. Stimulation intensity was adjusted, for each mouse,
1468 to ensure detection while minimizing heat dissipation in the cortex: this was achieved
1469 by operating the stimulation LED within a power range that produced an irradiance
1470 between 3.5mW/mm^2 and 11.5mW/mm^2. The refresh rate was set at 100Hz with a
1471 projected pixel size measuring approximately 6.5 μm. To prevent mice from directly
1472 perceiving the blue light used for optogenetic stimulations, we enclosed the optical
1473 column and the mouse skull with an opaque sleeve. To optogenetically stimulate the
1474 cortex with task relevant information concerning the vertical position of the stimulus,
1475 we shaped the blue light projection in every trial to match the top- or bottom- encoding
1476 subregions of the RL map. Light patterns were obtained by thresholding the average
1477 retinotopy movies within RL boundaries as defined by the fitted atlas. To effectively
1478 drive neuronal activity with blue light, we pulsed illumination at 30 Hz[63]. We used a
1479 pulsing pattern over the top or bottom encoding part of RL lasting one second within
1480 the response window. Before switching to the full optogenetic substitution task, mice
1481 that had achieved stable expert performance for three sessions in a row in the original
1482 tactile task first underwent 1 to 3 optogenetic habituation sessions. During these
1483 sessions, they were exposed solely to the Go light pattern (with a distribution of 80%
1484 Go trials and 20% Catch trials with light stimulations outside the cranial window). As
1485 soon as mice displayed a consistent licking response to optogenetic stimulations they
1486 were transitioned to the full task (with a distribution of 30% Go, 50% No go trials and
1487 20% Catch). Real-time projection control was achieved integrating ALP Vialux API
1488 with the custom-made Matlab graphical user interface (GUI) used to run the behavioral
1489 tasks and monitor mouse performance. Analysis of task performance during

optogenetic substitution sessions was carried out exactly as described in the "behavioral analysis" section above.

Network model

Code will be made available at: https://github.com/comp-neural-circuits. We constructed three recurrent excitatory (E) / inhibitory (I) rate networks representative of the areas V1, S1, and RL (Fig. 7a). We abstracted the decision process to a pair of readout units which in turn drove an integrator through a drift-diffusion process. Recurrent rates of the units in each region evolved according to

$$\tau_i \dot{r}_i = -r_i + \mathrm{f}\left(W_{ij} r_j + I(t) + \sigma_r \mathrm{x}_i(t)\right)$$

where W_{ij} are the connection weights from unit *j* to unit *i*, *I(t)* is a source of external input to V1 and S1 cells only, and x_i is noise independent to each unit, evolving according to $\mathrm{dx_i} = -\mathrm{x_i dt} + \sigma_x \mathrm{dW}$ where dW is a Wiener process with scaling factor σ_x. We used a saturating transfer function, given by:

$$\mathrm{f(x)} = \beta \left\lfloor \frac{1}{1 + \exp(-(\mathrm{x} - \mathrm{x}_0)/\rho)} - \frac{1}{1 + \exp(\mathrm{x}_0/\rho)} \right\rfloor_+ .$$

where $\lfloor \cdot \rfloor_+$ denotes a threshold at zero. Each region was divided into two subnetworks (corresponding to top- and bottom-preferring cells) of N_E excitatory neurons and $N_I = N_E$ inhibitory neurons each. These subnetworks were anatomically segregated such that weights were partitioned into top- and bottom-preferring units which did not interact (i.e. $W_{ij} = 0$ for *i* top and *j* bottom, or *i* bottom and *j* top; see Fig. 7). Within a subnetwork, neurons were all-to-all connected. Projections between regions respected subnetwork boundaries and were limited to E neurons. Weights of

feedforward projection connections (i.e. V1 or S1 to RL) were scaled slightly up relative to the within-region weights, and feedback projections (i.e. RL to V1 or S1) were scaled slightly down relative to within-region weights (see Table). To model experiments with TeNT-based silencing of RL we removed all incoming and outgoing connections to the model RL units. Simulations were subdivided into multiple trials of length 0.4 s, defined as a period of no external stimulation t_{isi} = 0.2 s, i.e. $I(t)$ = 0, together with presentation of a stimulus $I(t)$ > 0 to the V1 or S1 network (mimicking the visual or tactile tasks, respectively) for time t_{stim}= 0.2 s which was exclusive to either top- or bottom-preferring units in one of the two sensory regions (V1 or S1). The stimulus was modeled as a step function, i.e. $I(t) = c_s$, where c_s is a constant. The stimulus order was randomized for each simulation. In order to model rewards in our framework we considered two abstract readout units, z_1 and z_2, corresponding to Go and No go responses, respectively. The activity of these units was a weighted linear sum of activity from E neurons within the recurrent network $\mathrm{z_i(t)} = \alpha W_{ij} r_j$ where α is a scaling constant. The network's decision was determined by producing a decision variable δ_i for each z_i by integrating z_i from stimulus onset in the presence of independent noise:

$$\mathrm{d}\delta_i = z_i \mathrm{dt} + \sigma \mathrm{dW}$$

where dW is a Wiener process with scaling factor σ. Decisions were instantiated once a fixed threshold θ_0 was reached by either z_1 or z_2 resulting in the specific decision encoded by that readout unit; if neither z_1 nor z_2 crossed the threshold before the end of the stimulus period, the action was considered No go. Once a decision was reached the response z_i was modified according to $\tilde{z}_i(t) = z_i(t) + c_r$, until the end of the stimulus interval. If the decision was correct relative to the presented stimulus, then $c_r > 0$, otherwise $c_r < 0$. $|c_r|$ was defined as the difference between the current z_i and the last value of z_i before the stimulus period. Only projection weights from each

region (V1, S1, and RL) to the output variables z_i were plastic. Weights evolved according to a Hebbian rule:

$$\dot{W}_{ij} = \gamma z_i r_j$$

where $i \in \{1, 2\}$, $j \in \{1, \ldots, N_E\}$, and γ is the learning rate. Weights were bounded between 0 and $W_{max} = 1$ and normalized after each update to the initialized mean $\overline{W_z}$ within a region. Weights were initialized as a Gaussian centered at mean $\overline{W_z}$ = 0.5 with standard deviation $\sigma_z = 0.1$. All model simulations were run for 100 trials before and after a switch. To compare model performance with the animal behavior we recorded the trial outcome for each stimulus presentation. We then computed a performance metric and "action" probabilities conditioned on the stimulus in a sliding window of 15 trials. In particular, if we define s_1 as input to the bottom-preferring units and s_2 as input to the top-preferring units, within each window we calculated $P(z_1|s_1)$ and $P(z_1|s_2)$ as the conditional action probabilities. The performance accuracy was computed as P(correct) with correct defined as $z_1 > z_2 \mid s_1$ and $z_2 > z_1 \mid s_2$ in the rule-preserving switch condition and $z_1 > z_2 \mid s_2$ and $z_2 > z_1 \mid s_1$ in the rule-reversing switch condition. We quantified the number of trials to reach criterion across four different switch conditions: rule-preserving and rule-reversing in the full model; rule-preserving and rule-reversing with RL silenced. We set a criterion value of 75% performance accuracy for each trial condition to quantify the number of trials needed for the model to re-learn after a switch. Because the mean performance accuracy increased approximately monotonically while response conditions were held constant, we computed the first time at which a given model run reached the criterion value as the time to relearn. If a model run did not reach the criterion within 100 trials, we omitted it from this analysis.

Variable	Value
W_{EE}	0.16
W_{EI}	0.2
W_{IE}	0.16
W_{II}	0.18
c_s	30
ρ	15
x_0	20
β	30
α	$1/(6N_E)$
α (lesion)	$1/(2N_E)$
γ	4e-7
σ	0.5
σ_x	$\sqrt{2}$
σ_r	0.75
τ_E	20
τ_I	10
θ_0	80

1561

1562

1563 Quantification and statistical analysis

1564 Statistical details of experiments and analysis are described in figure legends and in
1565 the main text. Details include statistical tests used, sample type and size as well as
1566 definition of bar plots and error bars. In figure legends, standard error of the mean

(S.E.M.) is specified when plotted as error bars. Paired or unpaired t-tests were used to assess significance of mean comparisons (implemented by Matlab functions “ttest” and “ttest2”, respectively). Normality tests were not performed systematically but individual data points were plotted to visualize distributions. Wilcoxon signed rank test were used to assess significance in paired median comparisons (implemented by Matlab function “signrank”). Wilcoxon rank sum test was used for unpaired median comparisons (implemented by Matlab function “ranksum”). Pearson coefficient of correlation was used to compute correlations between two conditions (implemented by Matlab function “corr”). Across all fits reported uncertainties (i.e. confidence intervals) for best-fit parameter values were extracted from fit covariance matrices (fitting was performed using Matlab function “nlinfit”). In order to assess the statistical significance of multisensory modulation at each time bin, for each visuo-tactile stimulus condition, and for each responsive neuron, we performed a bootstrap-t procedure[64] with 2000 outer resamplings and 25 inner resamplings (over trials) in order to obtain p values for the difference from zero of MI values.

Data and Code Availability

The complete data set and Matlab analysis code will be made freely available at the CERN database, Zenodo.

Acknowledgements

We thank Denis Jabaudon, Alan Carleton, Daniel Huber and members of the El-Boustani laboratory for discussions. This work was supported by the Swiss National Science Foundation: PCEFP3_181070 (S.E.-B.).

Author Information

M.G., G.M. and S.E.-B. conceived experiments. M.G. performed surgeries, viral infections, wide-field and two-photon imaging, behavioral experiments, and post-mortem analysis. M.G., G.M. and C.G.F. performed behavioral experiments with optogenetics. G.M. coded acquisition software and performed data analysis. M.P.G. and J.G. conceived the network model and performed simulations. M.G., G.M. and S.E.-B. wrote the manuscript.

Ethics Declarations

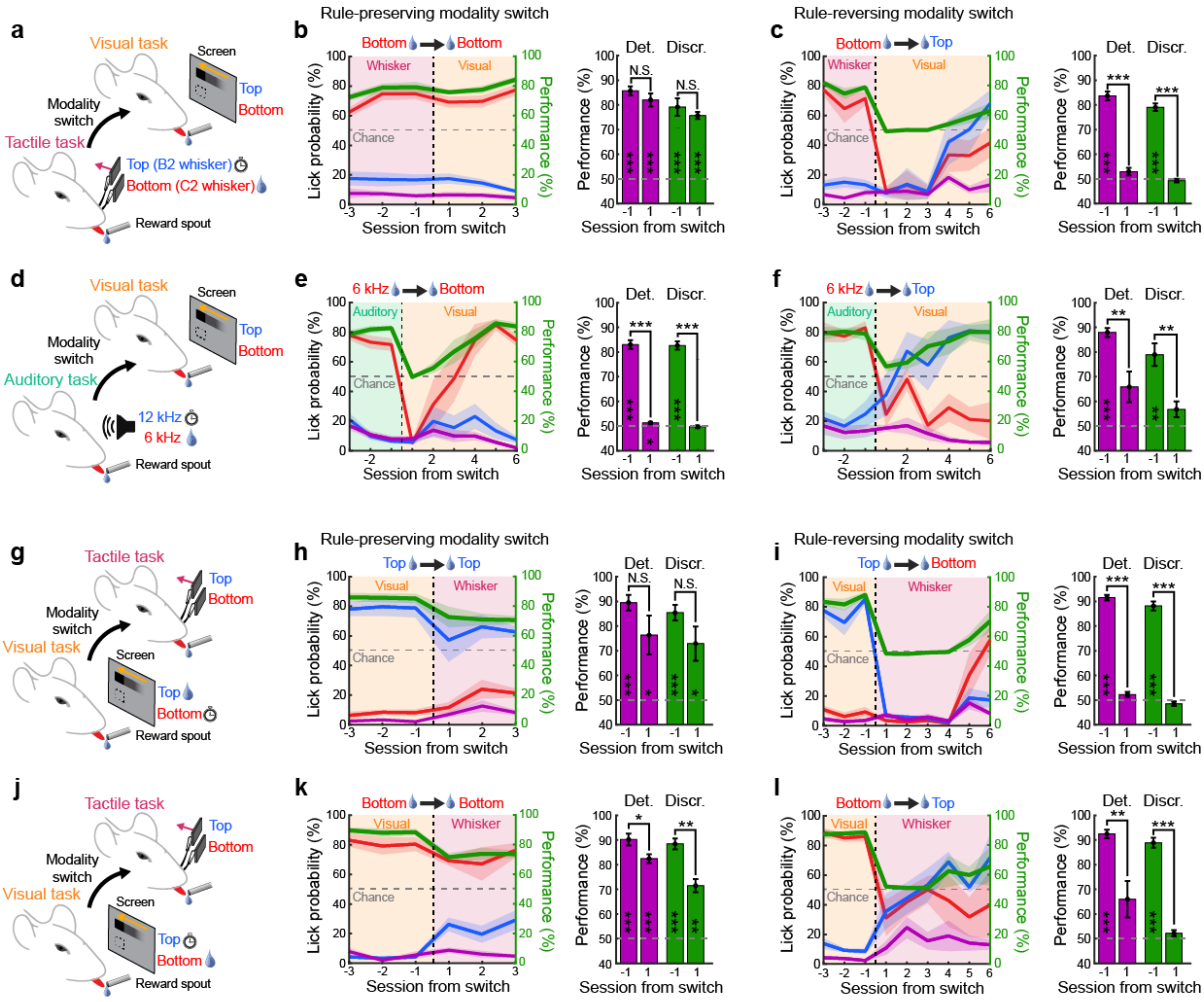
Authors declare no competing interests.

Corresponding author

Correspondence and requests for materials should be addressed to Sami El-Boustani (sami.el-boustani@unige.ch).

Supplementary Information

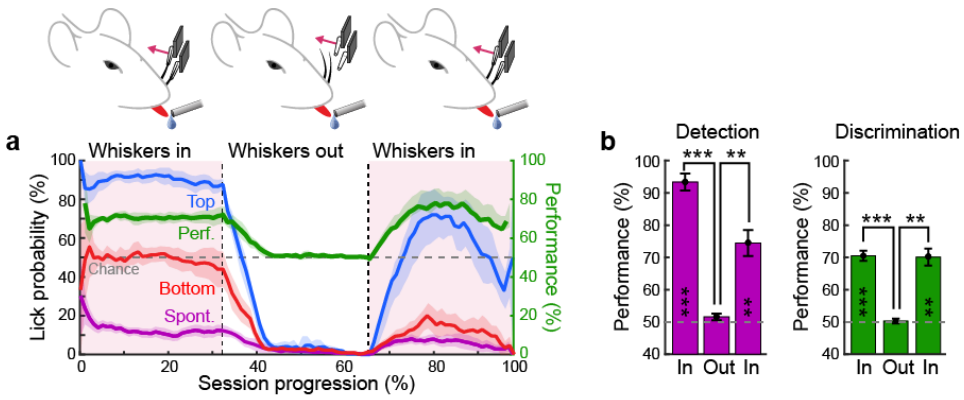
Extended Data Fig. 1-14



Extended Data Figure 1

Extended Data Fig. 1. Cross-modal generalization in mice across different conditions.

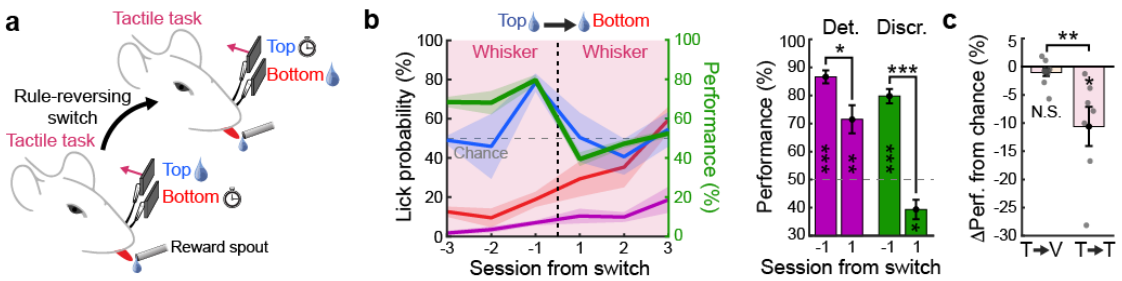
a, Schematic of behavioral paradigm where switches occur between a tactile task and a visual task. Bottom tactile stimuli were rewarded upon licking in the first task. **b,** Left: Average task performance and conditional lick probabilities across sessions for mice (N=5 mice) undergoing a rule-preserving modality switch (dashed vertical line). Shaded areas and color code as in Fig. 1e. Right: detection (purple) and discrimination (green) performance distribution the session before and after the switch (two-sided paired t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). **c,** Same as **b** but for a rule-reversing modality switch. **d,** Schematic of behavioral paradigm where switches occur between an auditory discrimination task with two pure tones (6 kHz and 12 kHz) and a visual task. The 6 kHz tone is always associated with a water reward. **e,** Same as **b** for an auditory to visual modality switch, where the bottom visual stimulus is associated to a reward. **f,** Same as **e** for an auditory to visual modality switch, where the top visual stimulus is associated to a reward. **g,** Schematic of behavioral paradigm where switches occur between a visual task and a tactile task with the top visual stimulus being the rewarded one. **h,** Same as **b** for mice undergoing a rule-preserving switch (N=5 mice). **i,** Same as **h** but for rule-reversing modality switches. **j,** Same as **g** with the bottom visual stimulus being the rewarded one. **k,** Same as **h** but for rule-preserving modality switch. **l,** Same as in **k** but for a rule-reversing modality switch. For cases with spatial rule-reversing modality switch or in absence of spatial prior (i.e. **c**,**e**,**f**,**i**,**l**), behavioral curves are shown for 6 consecutive sessions following the switch. We observed a small but significant drop in performance for switches from visual to tactile modalities under rule-preserving conditions (**h**,**k**) that was absent in the tactile to visual switches (**b**). This asymmetry might arise from generally higher initial performances in the visual task, the distinct nature of sensory representation (discrete for whiskers and continuous for visual inputs), or variations in circuit properties.



Extended Data Figure 2

Extended Data Fig. 2. Performance of mice at the whisker discrimination task with and without whiskers in the piezo stimulators.

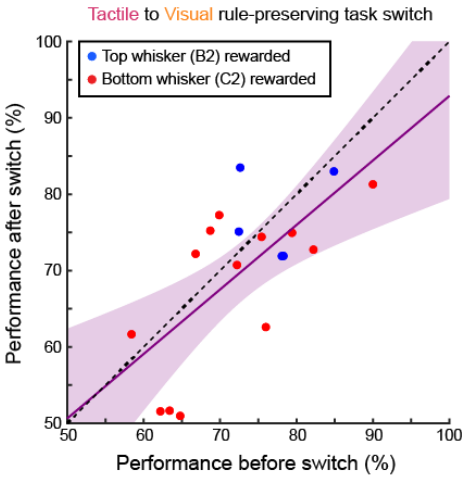
a, Conditional lick probabilities and performance during the whisker discrimination task averaged across expert mice (N=5 mice). Sessions were normalized between beginning (0%) and end (100%) before the average was performed. Color code is the same as in Fig. 1. During these sessions, mice performed the task for approximately one-third of the trials with whiskers present in the capillary tubes used to deliver tactile stimulations. Subsequently, the whiskers were removed while maintaining the tube positions, and the task continued for an equivalent duration. Finally, the whiskers were reinserted to confirm that mice were still motivated to perform the task. **b,** Comparison of detection (purple) and discrimination (green) performance during the 3 different phases of the sessions described in **a**. Comparisons are performed between consecutive phases (paired two-sided t-test, ***p<0.001, **p<0.01, *p<0.05) and with chance level (two-sided t-test with p=0.188 for detection and p=0.710 for discrimination during the Out phase). Mice were not able to perform the task in absence of direct whisker stimulations.



Extended Data Figure 3

Extended Data Fig. 3. Relearning the whisker discrimination task following a reversal of the spatial rule within the same modality.

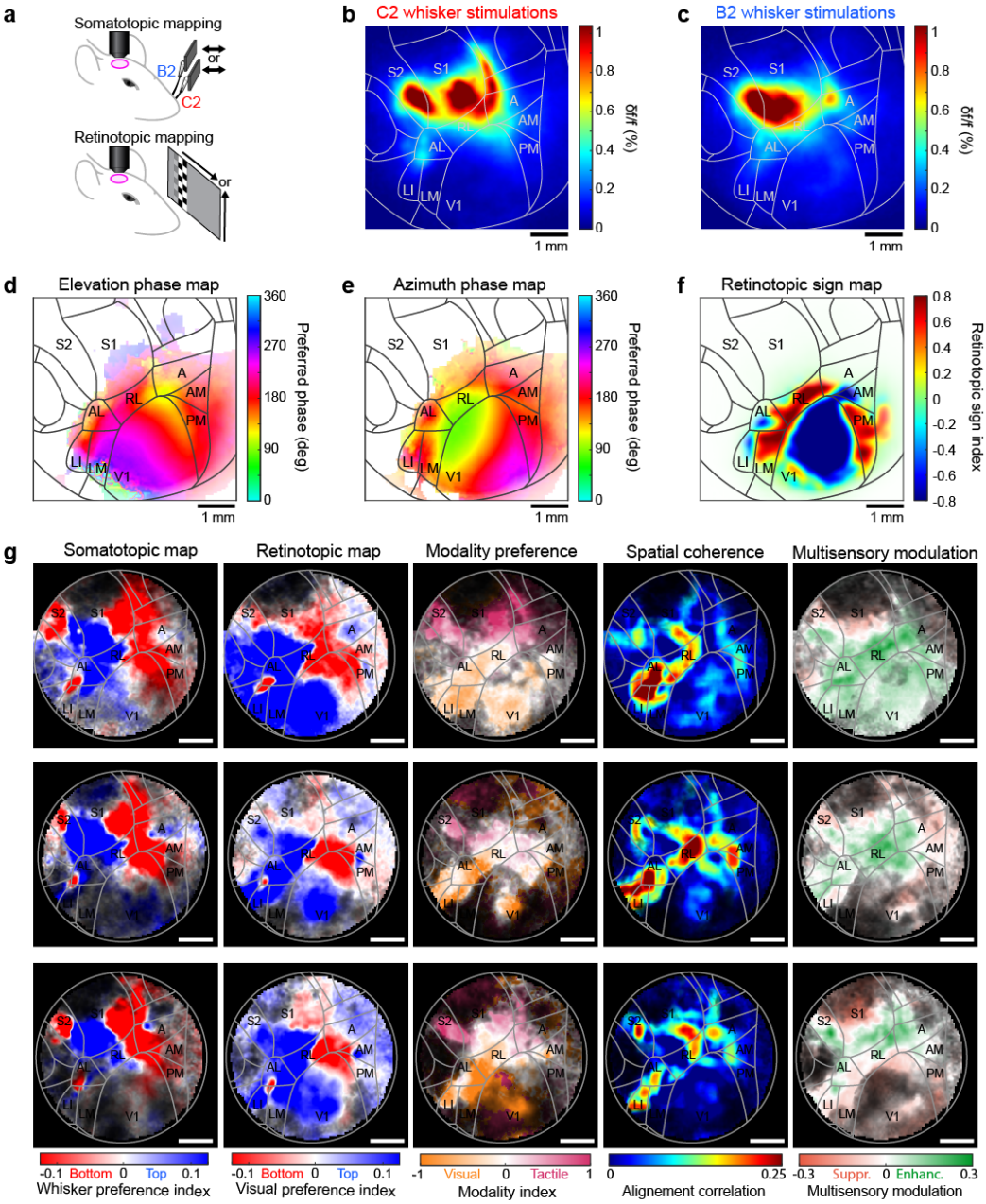
a, Schematic of the behavioral paradigm for rule-reversing switch within the tactile modality. Once mice became experts in the whisker discrimination task, the spatial rule of the task was reversed. **b,** Left: Average task performance across sessions for mice exposed to the reversed reward association (N=7 mice). Shaded area: S.E.M. Vertical dashed line indicates the transition from one rule to the other. Right: detection (purple) and discrimination (green) performance distribution the session before and after the switch (paired two-sided t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). **c,** Comparison of performance change relative to chance level after rule switch across sensory modalities (i.e. tactile to visual) or within the whisker task (N=10 mice for visuo-tactile and N=7 mice for tactile only, unpaired two-sided t-test, **p=0.0056). Mice maintained their responses to whisker stimulations significantly above chance level, yet their performance on the discrimination task fell significantly below chance (*p=0.0221). This suggests a reduced adaptability to a reversal of the spatial rule within the same modality, potentially due to a stronger habitual response linked to the previously rewarded whisker or an entrenched bias regarding the saliency of whisker stimulations.



Extended Data Figure 4

Extended Data Fig. 4. Performance level is maintained after switches from the whisker discrimination task to the visual task.

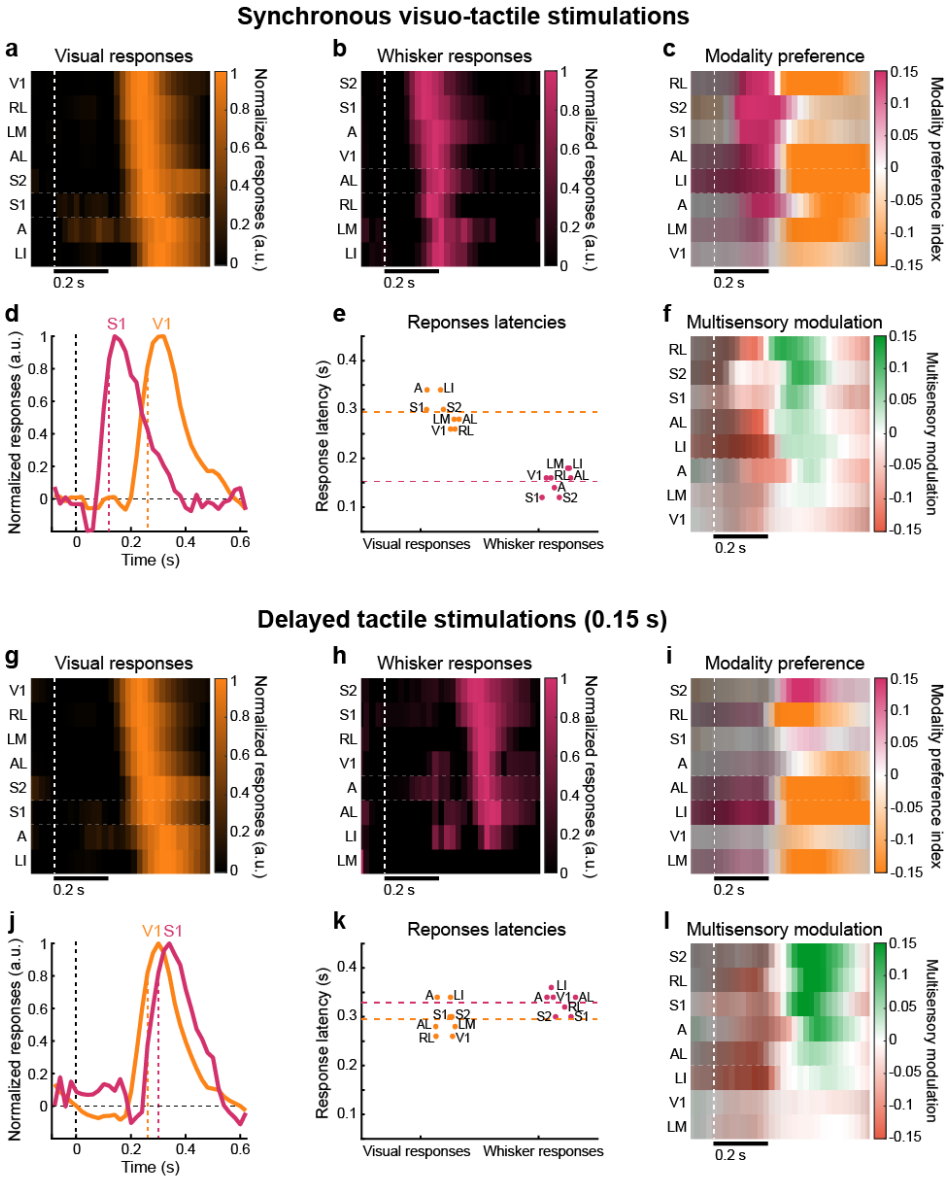
Comparison between performance in the whisker task the session before the switch and the performance in the visual task the session following the switch. This includes mice that experienced the switch, showcasing varying performance levels and rewarding either the top whisker (blue) or the bottom whisker (red). Performances before and after the switch were correlated (N=18 mice, Pearson coefficient: 0.68, p=0.0018) and close to the unity dashed line. Shaded area: 95% confidence interval. Mice exhibit consistent discrimination performance following modality switch, suggesting that cross-modal generalization may correspond to the initial sensorimotor learning level, or that shared influences, such as thirst-related motivation, impact both tasks equally.



Extended Data Figure 5

Extended Data Fig. 5. Average whisker and retinotopic maps with examples of single mouse visuo-tactile functional maps obtained in wide-field imaging.

a, Schematic of the protocols used to measure single whisker response maps (top) and retinotopic maps (bottom) in transgenic mice expressing GCaMP6f in layer 2/3 during wide-field imaging through a cranial window. **b,** Response map averaged across mice (N=41 mice) for C2 whisker stimulations. Projection of the Allen Mouse Brain Atlas is overlaid on top with areas names for reference. **c,** Same as in **b** for B2 whisker stimulations. **d,** Elevation phase map averaged over mice (N=50 mice). Pixels with low response amplitude are displayed in white. **e,** Azimuth phase map average over the same mice as in **d**. **f,** Retinotopic sign map computed based on the elevation and azimuth phase maps of **d** and **e**. **g,** Representative examples of visuo-tactile functional maps obtained from sparse noise protocol responses from 3 different mice (rows). From left to right: 1. Whisker preference map used to describe somatotopy in the vertical space with red colors indicating preference for the bottom whisker and blue colors indicating preference for the top whisker. 2. Visual position preference map used to describe retinotopy in vertical space. 3. Modality preference map describing the dominant modality for each pixel. 4. Spatial coherence map between retinotopic and somatotopic maps. 5. Multisensory modulation map indicating areas with responses enhancement or suppression for visuo-tactile stimuli.



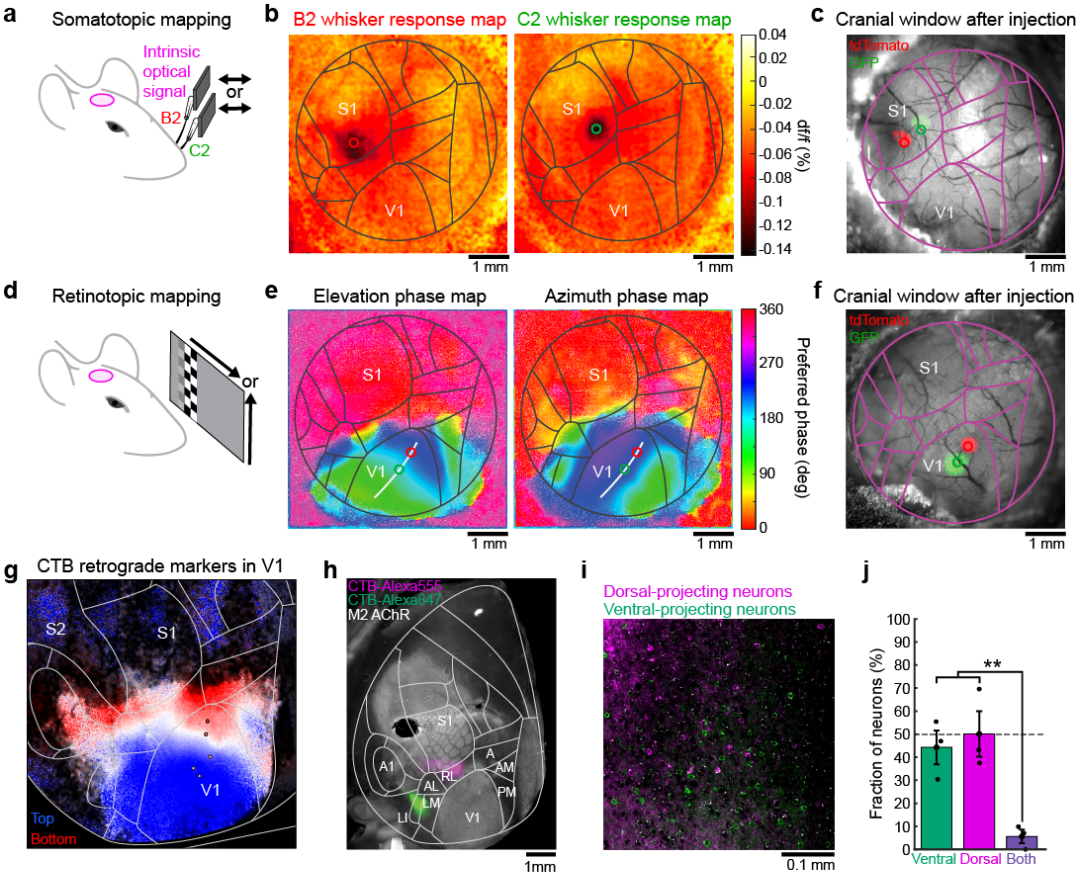
Extended Data Figure 6

1710

Extended Data Fig. 6. Comparison of cortical response dynamics for synchronous or delayed visuo-tactile stimulations.

a, Average activation time course for each visually responsive area following stimulus onset (N=29 mice). Areas are ordered based on response onsets. White dashed line indicates stimulus onset. **b,** Same as **a** for responses in cortical areas evoked by whisker stimuli. **c,** Time course of the modality preference index across cortical areas with visuo-tactile responses. **d,** Average response time course in the primary visual cortex (V1) and the whisker primary somatosensory cortex (S1) highlighting an offset of about 0.15 s between cortical responses evoked by the two modalities. Colored dashed lines indicate response latency computed as the time point where the response curve reaches 80% of its peak amplitude. **e,** Distribution of response latencies for cortical areas with evoked visual (orange) or whisker tactile (magenta) responses. Colored dashed lines indicate the average latencies across areas. **f,** Time course of the multisensory modulation index across cortical areas with visuo-tactile responses. **g-l** Same as **a-f** for tactile stimuli presented with a delay of 0.15 s with respect to the visual stimuli to match response latencies in the cortical areas. Multisensory enhancement was observed when stimuli were synchronized or delayed though they were stronger when cortical responses had comparable latencies.

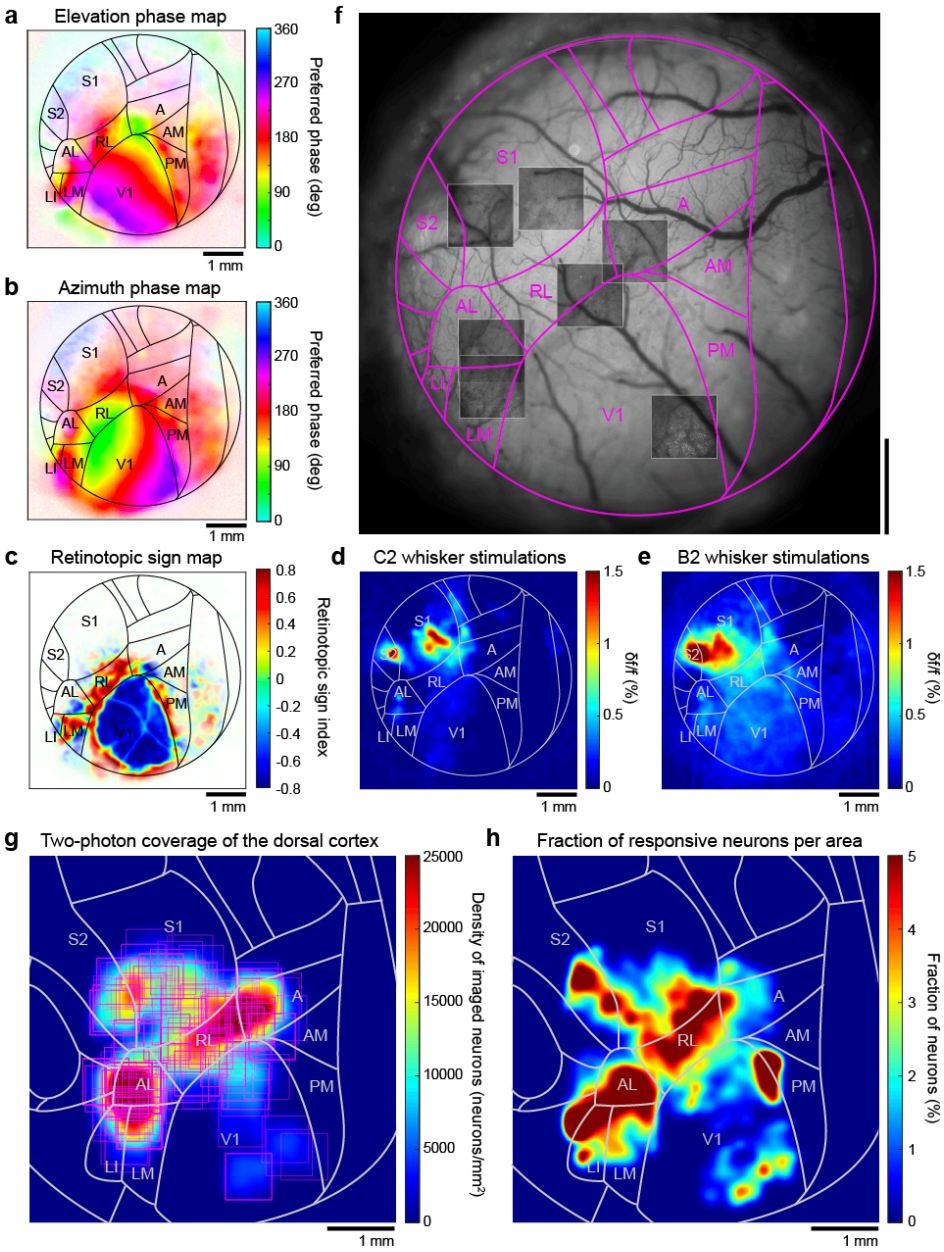
1729



Extended Data Figure 7

Extended Data Fig. 7. Functional mapping of vertical positions in the visual or whisker somatosensory cortex for anatomical tracing.

a, Schematic of the protocol for mapping whisker-responding areas in the dorsal cortex. Single whiskers are stimulated repeatedly during intrinsic optical signal imaging. **b,** Change in red light reflectance on the surface of the cortex measured after averaging several trials of whisker stimulations relative to baseline. The left image reveals the barrel location corresponding to B2 whisker stimulations and the right plot for C2 whisker stimulations. **c,** Image of the cranial window several weeks after injection of viral vectors to express GFP and tdTomato in two distinct parts of the whisker somatotopic map. Locations of B2 and C2 barrels in S1 defined in **b** are indicated with circles overlaid on top of the fluorescent signal where injections were performed. **d,** Schematic of the protocol for retinotopic mapping in areas of the dorsal cortex. Horizontal or vertical bars drift across the screen and they are textured with a flickering checkerboard pattern to elicit strong cortical responses at different phases corresponding to the retinotopic preference. **e,** Retinotopic phase maps for elevation and azimuth on the dorsal cortex. These maps are then used to identify a gradient across vertical dimension in an iso-profile along the azimuth dimension. Two retinotopically distinct locations along the vertical axis are identified and labeled for viral injection. **f,** Image of the cranial window four weeks after injection of viral vectors to express GFP and tdTomato in two distinct parts of the primary visual cortex retinotopic map. Locations where stereotaxic injections were performed corresponding to **e** are indicated with circles overlaid on top of the fluorescent signals. **g,** Cholera Toxin B subunit (CTB) conjugated to different Alexa dyes was injected in different locations of the retinotopic map in V1 (circles). After perfusing the brain and flattening the cortex, a map was reconstructed over the dorsal cortex combining all injection sites to reconstruct areas where neurons project to the region of V1 encoding stimuli on the top (blue) or bottom (red) part of the visual field. All cortices with CTB labels were registered to a common atlas before performing the analysis (5 injection sites from N=3 mice). A similar analysis was performed previously with coronal slices[29]. **h,** Flattened cortex with injections of CTB conjugated with two different dyes (Alexa555 and Alexa647) in the dorsal and ventral visuo-tactile area, respectively. Immunohistochemical localization of M2 AChR is used to identify V1 and the barrels in S1 for atlas registration. **I,** Example image obtained with confocal microscopy where dorsal-projecting neurons (magenta) and ventral-projecting neurons (green) are intermingled in V1. **j,** Fraction of neurons that projected to the ventral area exclusively (green), to the dorsal area exclusively (magenta) or to both (purple) (N=3 mice, unpaired two-sided t-test, $^{**}p=1.8\times10^{-3}$, n=230 neurons for ventral, n=277 neurons for dorsal and n=26 neurons for both out of 26 fields-of-view). Neurons that project to the dorsal and ventral visuo-tactile areas display minimal overlap, suggesting they may process distinct types of visual information.

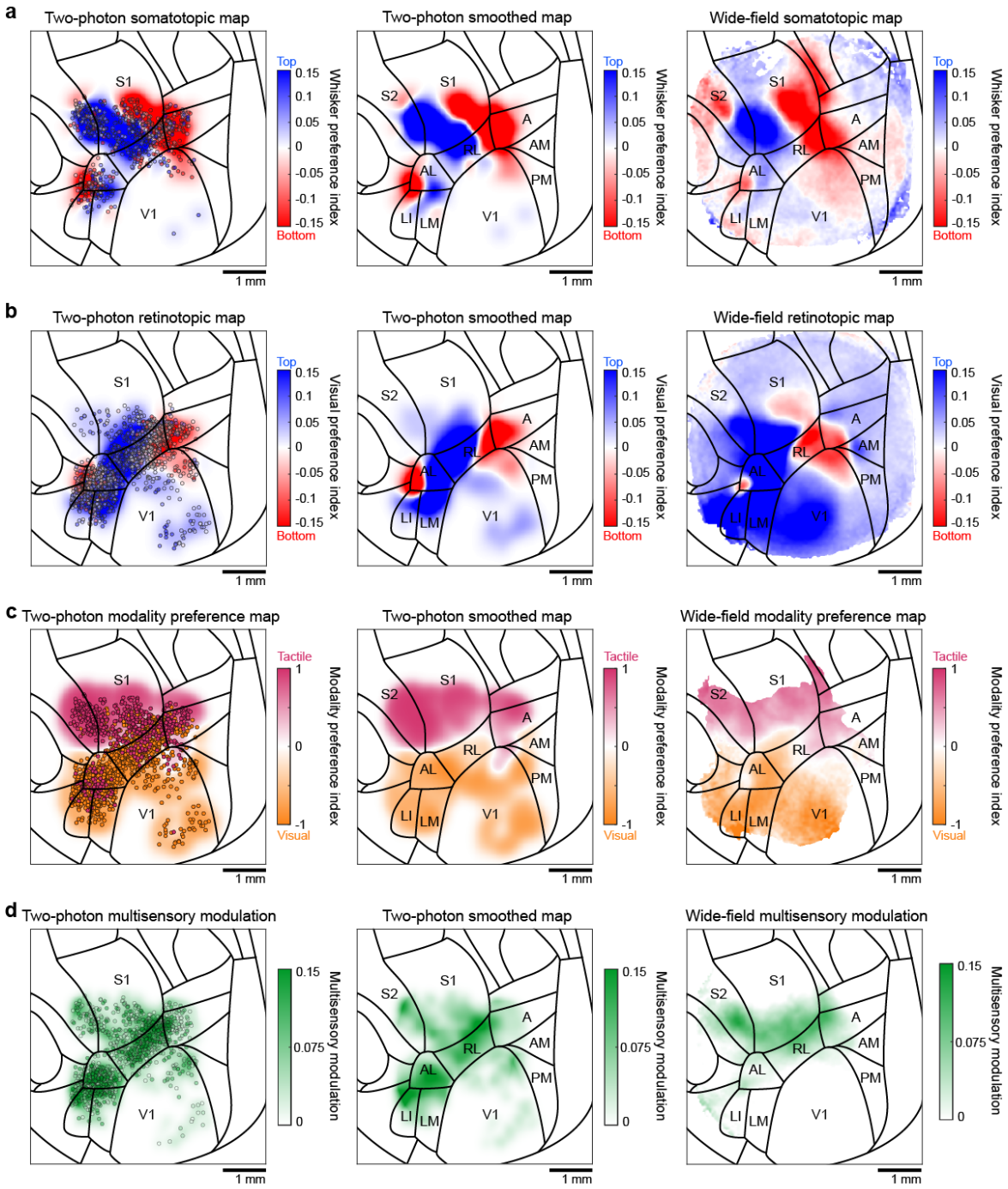


Extended Data Figure 8

1767

Extended Data Fig. 8. Coverage of large portions of the cranial window for characterizing visuo-tactile responses at single-cell level using two-photon microscopy.

a-e, Functional maps are obtained and aligned with the Allen Mouse Brain Atlas for each individual mouse, ensuring precise localization of neurons within their respective areas. Subsequently, all maps are realigned to a common reference atlas to reconstruct wide-field maps. Visual stimulation protocols are used to obtain phase maps for elevation (**a**) and azimuth (**b**) to compute the retinotopic sign map (**c**). Additionally, whisker stimulations provide response maps for C2 (**d**) and B2 (**e**) whisker stimulations. **f,** Image obtained with wide-field imaging where blood vessels are well distinguished on the surface of the cortex. The atlas fitted with functional maps is overlaid on top in purple. For this mouse, seven fields-of-view were imaged using two-photon microscopy and positioned on the cortical surface by matching blood vessel patterns. **g,** Coverage of visual and tactile areas of the dorsal cortex with two-photon imaging across mice. Using the common atlas and information about all neuron's location, we can recreate a large-scale population using single-cell information. We could collect N=272 fields-of-view across N=25 mice. **h,** Fraction of neurons imaged with two-photon microscopy that were responsive to at least one stimulus condition. The areas of high responsivity are in accordance with wide-field imaging.



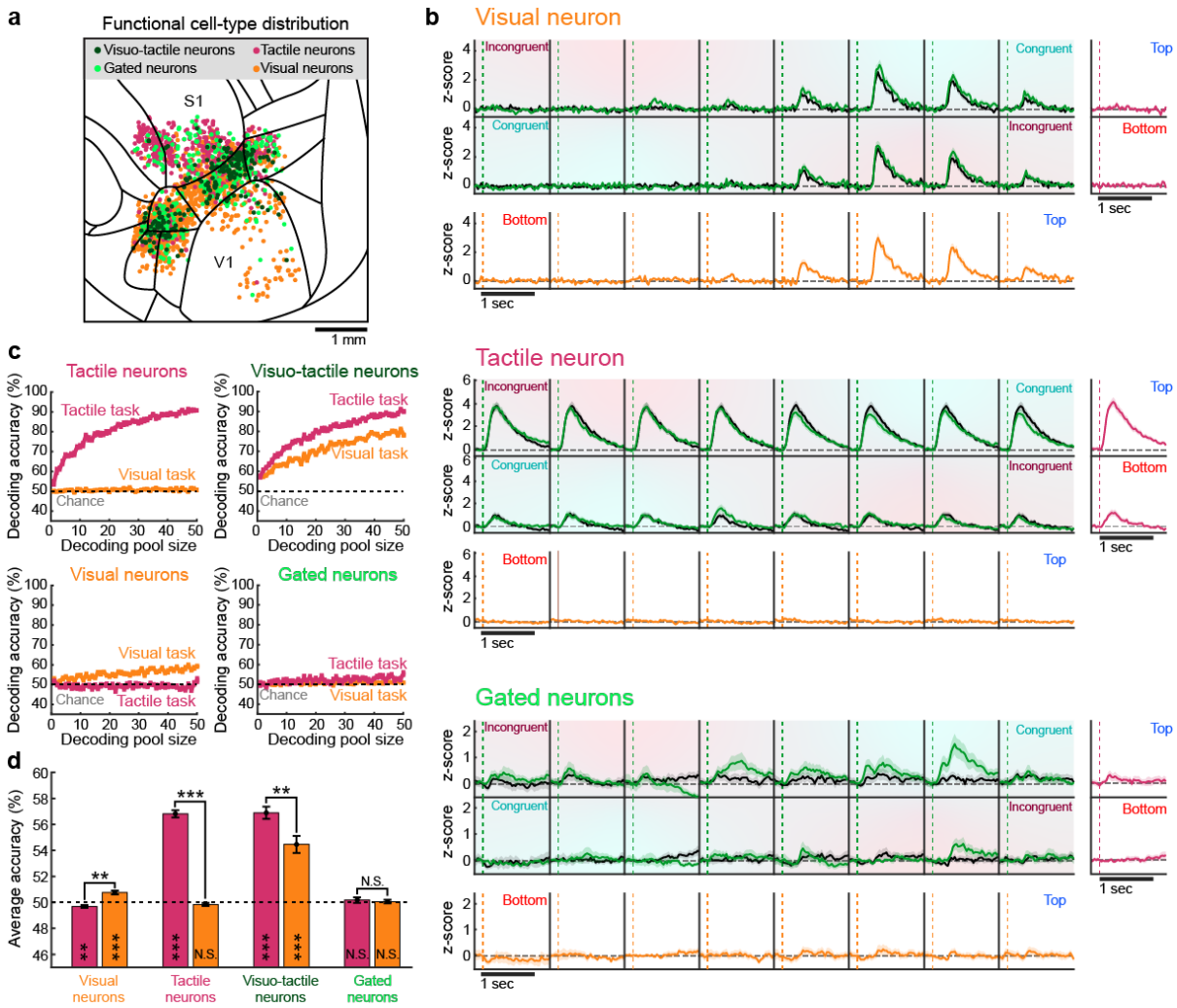
Extended Data Figure 9

1786

Extended Data Fig. 9. Functional maps correspondence between wide-field and two-photon calcium imaging.

a, Left: Distribution of neurons imaged with two-photon microscopy with significant whisker tactile responses. Neurons are displayed on the common atlas and are color-coded based on their whisker preference (n=790 neurons). Middle: Large-scale functional map reconstructed from single neuron whisker preferences. Right: Somatotopic map for the whisker preference obtained with wide-field imaging. Pearson coefficient of correlation between the wide-field map and the two-photon reconstructed map: 0.847. **b,** Same as in **a** for neurons with significant visual responses (n=1,816 neurons, Pearson coefficient of correlation: 0.651). **c,** Same as in **a** displaying modality preference for all neurons with significant whisker tactile or visual responses (n=2,160 neurons, Pearson coefficient of correlation: 0.795). **d,** Same as in **a** displaying multisensory enhancement for all neurons with significant multisensory modulation. For this analysis, the multisensory modulation is computed as the variation between the most suppressed and most enhanced response across conditions for all neurons (n=2,497 neurons, Pearson coefficient of correlation: 0.4458).

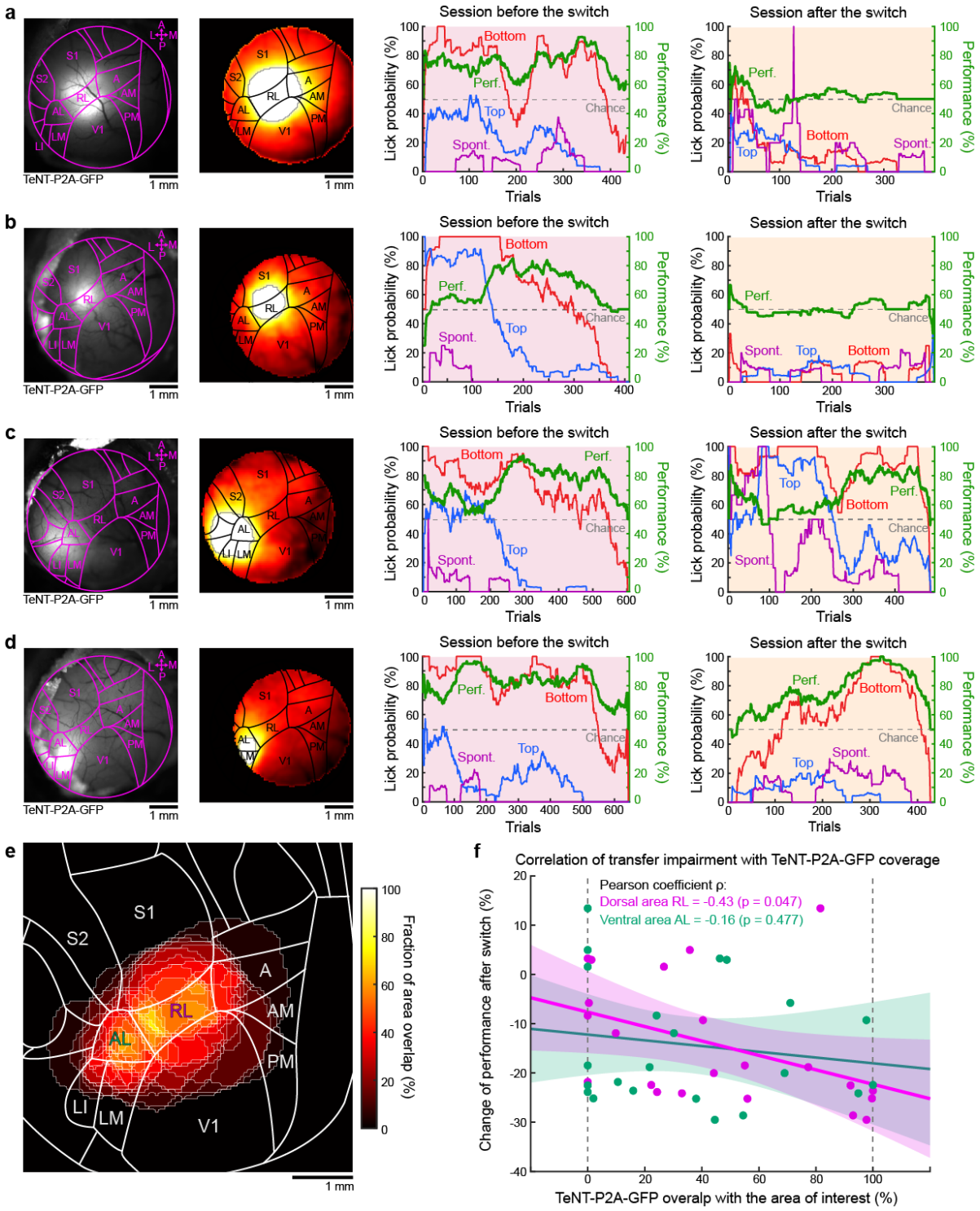
1803



Extended Data Figure 10

Extended Data Fig. 10. Distribution of functional cell-types across the dorsal cortex and their impact on cross-modal generalization.

a, Distribution of significantly responsive neurons across the dorsal surface of the cortex (from N=25 mice). Common atlas overlaid for reference. Color code indicates functional cell-type of each neuron (n=2,703 neurons). Neurons were classified in four categories: tactile neurons (magenta) responded only to whisker stimulations (n=567); visual neurons (orange) responded only to visual stimulations (n=1,593); visuo-tactile neurons (dark green) responded to both visual and whisker stimulations (n=213); gated neurons (light green) responded only to concomitant visuo-tactile stimulations (n=319). **b,** Example visuo-tactile receptive field for a visual neuron (top), a tactile neuron (middle) and a gated neuron (bottom). Color code is identical to Fig. 4b. Note that the gated neuron responds only when visual and whisker stimulations are presented together in the top location (spatially congruent condition). **c,** Decoding accuracy for the tactile task and the visual task as a function of the pool size. A Bayesian decoder was trained to discriminate the location of whisker stimulation (magenta) based on the neuronal responses of a pool of neurons. The same decoder was then used to decode the location of visual stimuli (orange), mimicking the modality switch experienced by mice in the behavioral experiment. Error bars: S.E.M. Only visuo-tactile neurons were able to achieve good generalization performance across modalities because of their functional properties. Tactile neurons could only afford good decoding performance in the whisker-based task but could not maintain it when switching to the visual task. Performance chance level is indicated with a gray dashed line at 50%. **d,** Comparison of single-neuron decoding accuracy distributions on tactile responses (magenta) and generalization to the visual ones (orange) for each functional cell-type class (paired two-sided t-test comparing tactile and visual task, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Decoding accuracies are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, N.S.: Not significant). Error bars: S.E.M.

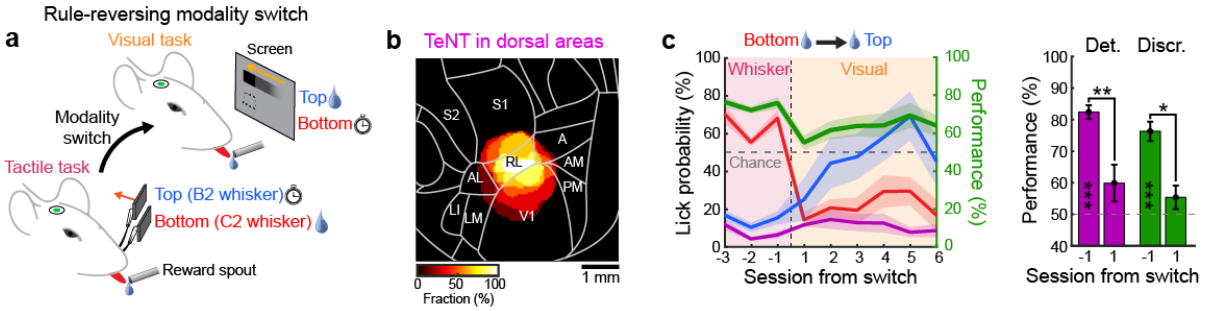


Extended Data Figure 11

1832

Extended Data Fig. 11. Cross-modal generalization impairment following rule-preserving switches in relation to TeNT expression in different associative areas of the dorsal cortex.

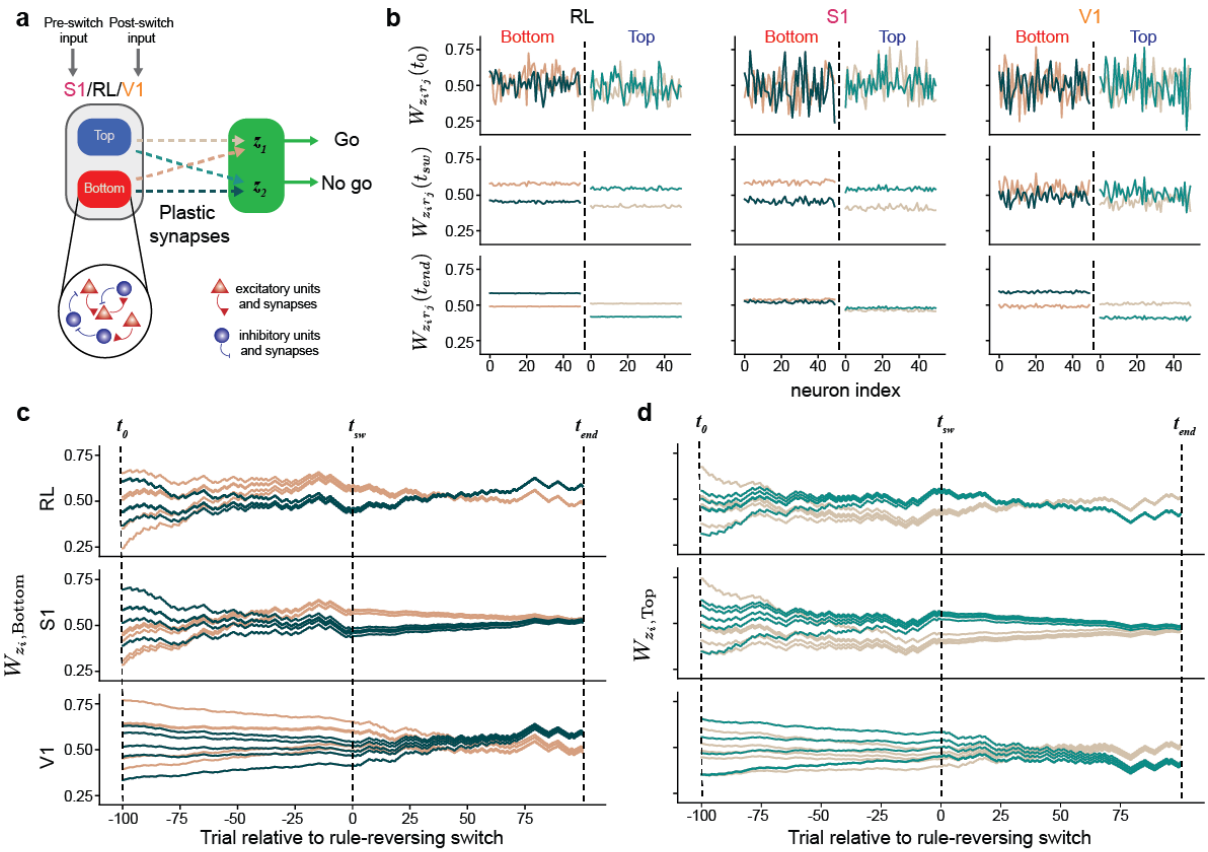
a, Left: Example of TeNT-P2A-GFP expression pattern over the cranial window and corresponding estimate of expression coverage after blood vessels subtraction. Right: Effect of cortical silencing on cross-modal generalization from the tactile task to the visual task while preserving the spatial rule of the task. The discrimination performance dropped at chance level after the switch. **b-d,** Same as in **a** but for other example mice with cortical injections at various locations and the resulting impact on cross-modal generalization performance. Note that equal volumes of viral vectors were injected in each mouse; however, some expression patterns appear truncated at the edge of the window. **e,** Overall coverage across all experiments performed with this protocol. **f,** Example of correlation between the level of TeNT-P2A-GFP overlaps with a specific area (here RL or AL) and the corresponding drop of discrimination performance after switch. To obtain a more robust effect for the correlation, we computed the performance changes over 3 consecutive sessions after switch (N=22 mice, Pearson coefficient: -0.43, p=0.047 for RL; Pearson coefficient: -0.16, p=0.48 for AL).



Extended Data Figure 12

Extended Data Fig. 12. Relearning after rule-reversing modality switch in mice expressing TeNT-P2A-GFP in the dorsal area RL.

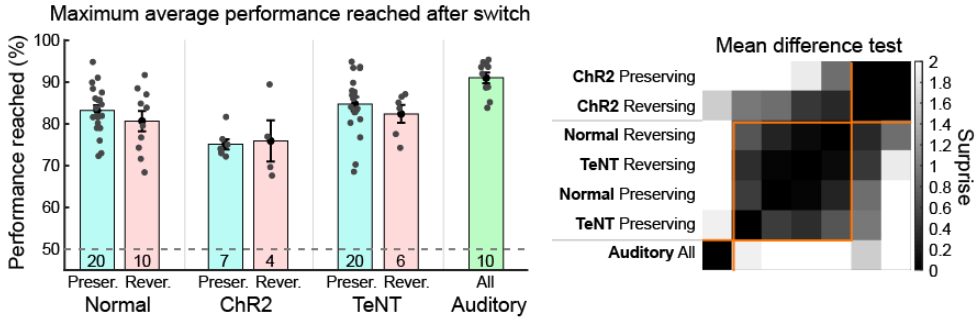
a, Schematic of behavioral paradigm where the modality switches occur between a tactile task and a visual task in mice expressing TeNT-P2A-GFP in RL. The bottom whisker stimulus is associated with the reward initially while the top visual stimulus becomes associated to rewards after the modality switch. **b,** Average TeNT-P2A-GFP coverage in all mice where dorsal neurons were silenced (N=6 mice). **c,** Left: Average task performance and conditional lick probabilities across sessions for mice population in **b** (N=6 mice) switching from the tactile task to the visual task with a spatial rule reversal. Shaded area: S.E.M. Black dashed line indicates the switch between modalities. Right: detection (purple) and discrimination (green) performance distribution the session before and after the switch (paired two-sided t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). Contrary to mice with intact cortical activity, mice with silenced RL area relearned the visual task faster after the rule-reversing modality switch, indicating a lack of prior on the spatial rule in these mice despite their previous experience with the tactile task.



Extended Data Figure 13

Extended Data Fig. 13. Model performance on a rule-reversing modality switch.

a, Model schematic. The model consists of three networks for areas S1, V1, and RL (here only one network is shown for simplicity) that comprised excitatory and inhibitory units tuned to either bottom or top stimuli. Dashed lines denote plastic synapses, and their colors are consistent throughout the figure. The excitatory units in all networks project to the two decision variables z_1 and z_2. S1 was initially stimulated, and Bottom stimuli mapped to "Go" action. After the switch, V1 was stimulated, and Top stimuli were mapped to "Go" action. **b,** Weights shown for different time points (corresponding to the dashed lines in panels **c** and **d** as a function of the neuronal index. $W_{z_i,r_j}(t)$ denotes the weights from sensory neuron j to decision variable i at time t. t_0: initial timepoint; t_{sw}: time of the switch; t_{end}: final timepoint after the switch. At the time of the switch (t_{sw}), all regions have the same organization of weights such that the correct output is produced (including in V1, due to the reverberation of network dynamics). After the switch, areas RL and V1 have reorganized to drive the correct output. **c,** Subsample of weights from the Bottom-preferring units to z_1 (brown) and z_2 (teal) as a function of time in a simulation. **d,** Same as in **c** for Top-preferring units.



Extended Data Figure 14

Extended Data Fig. 14. Maximum average performance reached following modality switch across different conditions.

Left: Maximum average performance reached in the discrimination task following the modality switch. Mice numbers are indicated for each group at the bottom of the bar. Mice that did not reach the expert criteria in Figure 7F are discarded from this plot. Error bars: S.E.M. Right: Surprise matrix computed from pairwise unpaired two-sided t-test between conditions. A hierarchical clustering based on cosine similarity was used to group conditions based on surprise values. Mice were clustered with respect to the maximum performance reached based on the nature of sensory stimulation used after switch (visual, auditory or optogenetic).