Contextual signals in visual cortex

How sounds, state, and task setting shape how we see

oude Lohuis, M.N.

Publication date
2022

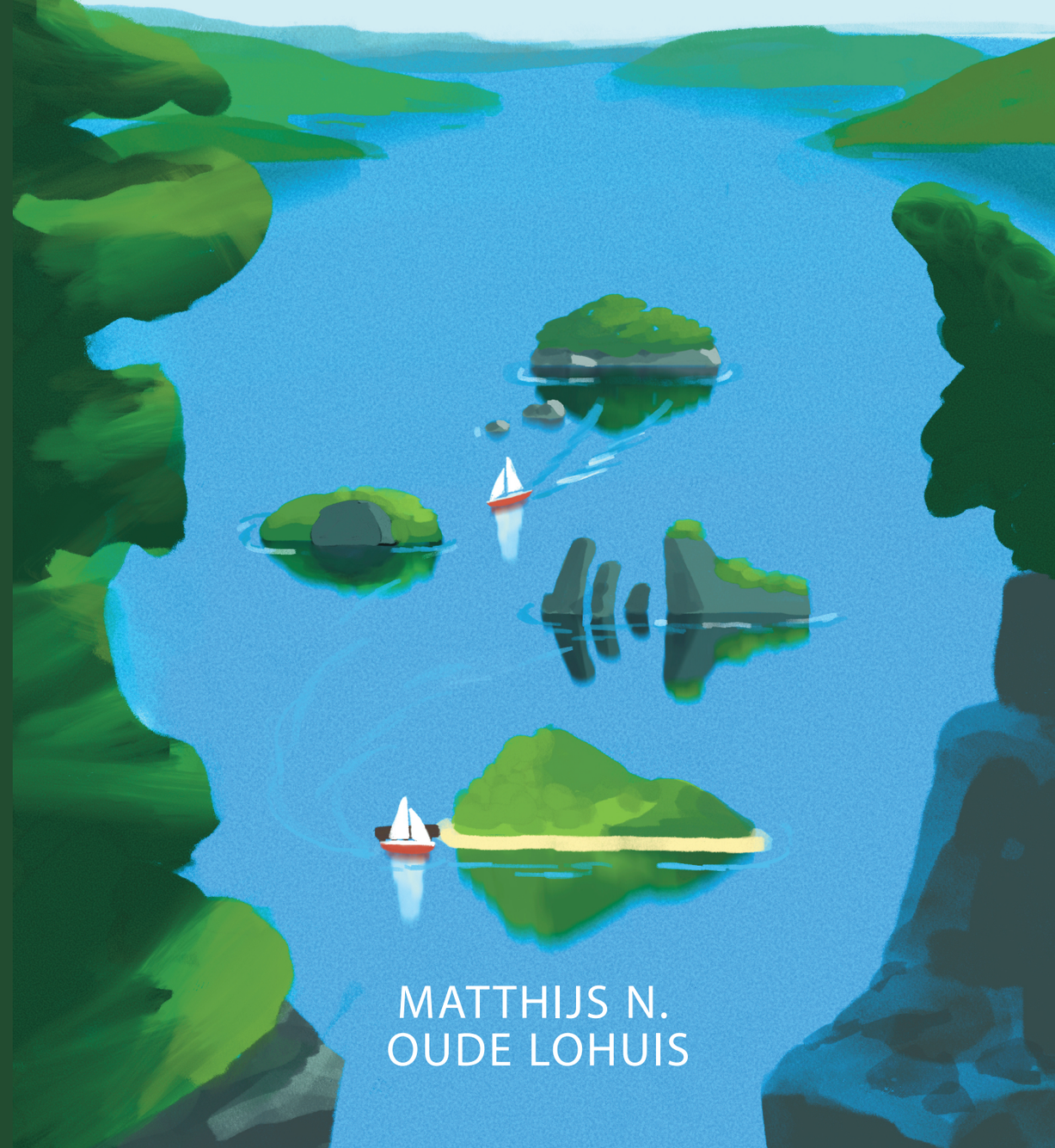
Document Version
Final published version

CONTEXTUAL SIGNALS IN VISUAL CORTEX

MATTHIJS N. OUDE LOHUIS

CONTEXTUAL SIGNALS IN VISUAL CORTEX

HOW SOUNDS, STATE, AND TASK SETTING SHAPE HOW WE SEE



MATTHIJS N.
OUDE LOHUIS

Contextual signals in visual cortex

how sounds, state, and task setting shape how we see

Matthijs N. Oude Lohuis

The research described in this thesis was carried out in the research group Cognitive and Systems Neuroscience at the Center for Neuroscience of the Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam, The Netherlands.

Cover artwork by Diogo Matias. Contextual vision is symbolized as interactions between brain areas with information travelling from islands of sound (speaker), vision (eye) and cognition (thought cloud).

ISBN/EAN: 978-94-6469-011-8
Printed by: ProefschriftMaken || www.proefschriftmaken.nl

Contextual signals in visual cortex

How sounds, state, and task setting shape how we see

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 6 oktober 2022, te 16.00 uur

door Matthijs Nicolai oude Lohuis
geboren te Amsterdam

Promotiecommissie

Promotor:	prof. dr. C.M.A. Pennartz	Universiteit van Amsterdam
Copromotor:	dr. U. Olcese	Universiteit van Amsterdam
Overige leden:	prof. dr. S.M. Bohte	CWI
	prof. dr. H.W.H.G. Kessels	Universiteit van Amsterdam
	dr. S. van Gaal	Universiteit van Amsterdam
	dr. J.F. Mejias	Universiteit van Amsterdam
	prof. dr. U. Noppeney	Radboud Universiteit Nijmegen
	prof. dr. J.K. Bizley	University College London
	prof. dr. C.N. Levelt	Vrije Universiteit Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

It is foolish to investigate sensory mechanisms blindly—one must also look at the ways in which animals make use of their senses. It would be surprising if the use to which they are put was not reflected in the design of the sense organs and their nervous pathways as surprising as it would be for a bird's wing to be like a horse's hoof.

Horace Barlow, 1961

We sometimes call these electric potentials 'messages', but we have then to bear in mind that they are not messages in the sense of meaningful symbols. To call them signals presupposes an interpreter, but there is nothing to read 'signals' any more than 'messages'. The signals travel simply by disturbing electrically the next piece on their route.

Sir Charles Sherrington, 1941

Contents

Chapter 1 : General Introduction

The inner workings of the brain pose perhaps the most baffling problem in science. Throughout millions of years of evolution, this organ has developed immense spatiotemporal complexity. Brief electrical potentials are communicated in a web of countless intermingled synapses. Billions of specialized neurons collectively sample the environment, integrate recent with previous experience, model the state of the external world, and control and guide adaptive behavior.

To follow what is happening in the external world (but also the internal world, i.e. in the body) the nervous system uses an exquisite variety of sensory systems. Among these senses, vision is by far the dominant modality for us humans, and vision is so inseparable from our day-to-day understanding of the world that we tend to say that '*seeing is believing*'. For this reason - and practical methodological motivations - the visual system has been especially intensively studied. Seminal work in anesthetized cats found that repeated presentation of well-defined local visual stimuli evoked stereotypic neuronal responses in the primary visual cortex (Hubel and Wiesel, 1959). Further work in immobilized and anesthetized preparations gave rise to the view that single neurons act as passive feature detectors, each neuron firing when a particular element of the outside world is present (Barlow, 1961). The detection of a total set of visual features leads to a representation of a visual scene: with the same visual input the same scene would be reconstructed.

However, there is a growing understanding that the way we perceive the world is highly contextual and depends on many additional factors beyond the photons that hit our retina. For example, we fail to perceive a gorilla walking through a scene if we are focusing our attention on counting passes between basketball players (Simons and Chabris, 1999). Moreover, sounds can make us falsely perceive flashes that are not there (Shams et al., 2000) or change how we interpret lip movements during speech (Mcgurk and Macdonald, 1976). A car might be spotted when driving to work after you had your morning coffee, but not on the way back after a long day of work. In short, what we perceive is strongly influenced by what we are focusing on, what we expect to see, what else is happening, what our goal is, and what our state is: are we drowsy, highly alert, or anesthetized?

So, even though the underlying circuitry remains principally unchanged, visual processing displays a large range of operating modes across conditions and environments. This ability to change behavioral and neural responses to identical sensory stimuli depending on the current task and context is an essential component of flexible goal-directed behavior. It is exactly this capacity to flexibly respond to changing contexts that is likely attributable to the cortex (Freedman and Assad, 2006; Mante et al., 2013; Miller and Cohen, 2001) and holds not only for vision, but all sensory modalities (Condylis et al., 2020; Sakata, 2016).

From vision to action

In this introduction, I will briefly discuss prevalent ideas about how incoming visual information can be processed from the retina to frontal motor areas to guide behavior in the external world, i.e. *perceptual decision making*. However, this process is flexible and dependent on multiple contextual factors. I will outline how three key processes dynamically modulate visual processing (state, sound, and task setting) and highlight anatomical and functional motifs of how this might be implemented. Key organizational principles of the visual system were established in primates and felines. However, in view

of the powerful array of experimental tools available for mice, the focus of this introduction will be on how visually-guided action depends on contextual factors in this species. Note that *context* in this thesis is meant in a broad sense, namely as factors beyond the locally available visual input, and thus operationalized as modulation by sources external to visual cortex, sources that convey information other than that of retinal origin.

Primary visual cortex

The visual pathway of mammals starts at the retina, where retinal ganglion cells project to the dorsal lateral geniculate nucleus of the thalamus (dLGN), as well as several other subcortical regions, including the superior colliculus and suprachiasmatic nucleus. The dLGN in turn projects to the primary visual cortex (V1) and forms the largest pathway for visual information to reach the cortex. As in all mammals, these projections are highly structured and follow a retinotopic organization (Kaas, 1980). This entails that the spatial organization of light as it hits the retina is conserved throughout the first stages of the visual system. Therefore, the visual cortex reflects a 'map' of the visual field whereby stimuli coming from the same part of the visual field are also represented by neighboring V1 neurons (Hübener, 2003; Smith and Häusser, 2010) (Figure 1.1). Single neurons in V1 thus process local visual inputs from a region in retinal space (their receptive field) and from the pioneering work by David Hubel and Torsten Wiesel in cats we know that they preferentially respond to contrasts along a certain orientation (Hubel and Wiesel, 1959, 1968). A neuron's receptive field can often be approximated well by a so-called Gabor patch, a sinusoidal grating convolved with a Gaussian envelope (Mehrotra et al., 1992). To compose visual inputs using these local filters has been argued to result from an optimization process to efficiently (en)code visual scenes; in other words, to represent a particular stimulus a minimal amount of action potentials is needed across a population of V1 neurons (Olshausen and Field, 1997; Tolhurst et al., 1981). Visual cortical neurons thus respond to locally oriented contrasts ('edges') and enhance the signaling of contours and edges relative to nondescript isoluminant surfaces (Marr and Hildreth, 1980). A dominant approach in the study of cortical circuits has therefore been to use orientated contrasts (gratings) that reliably and predictably evoke activity in V1 neurons.

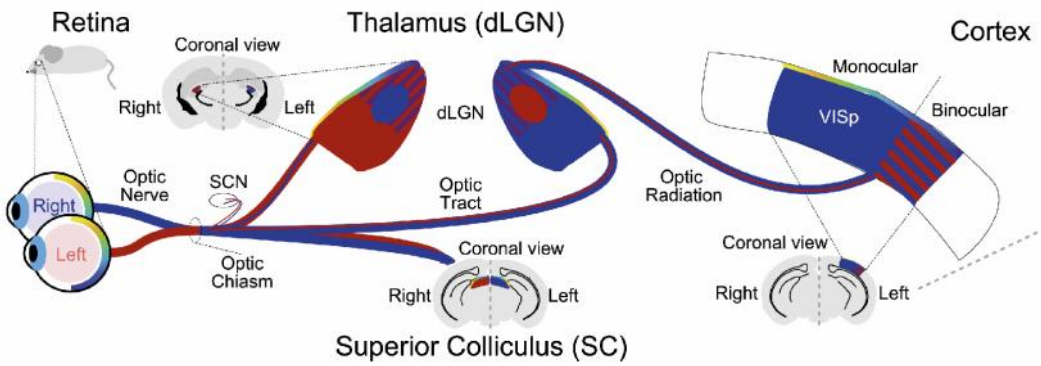


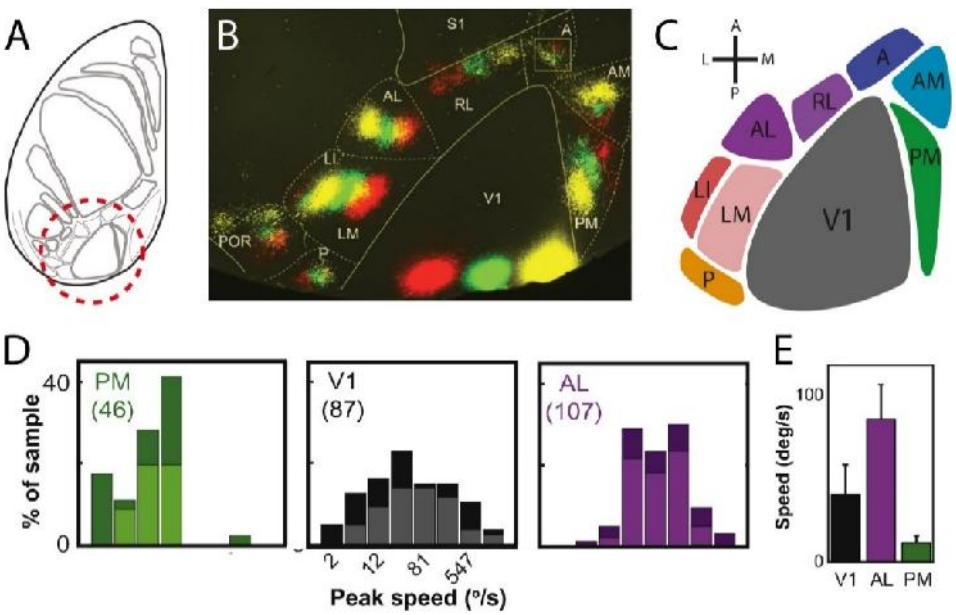
Figure 1.1: Pathway of the mouse visual system. *Visual information enters the brain via the retina. Output from the eyes is shown as red (left eye) and blue (right eye). Output via the optic nerve projects to targets in the suprachiasmatic nucleus in the hypothalamus (SCN), superior colliculus in the midbrain (SC), and dorsolateral geniculate nucleus of the thalamus (dLGN). Insets show SC and dLGN in coronal brain sections. Visual information crosses the midline to represent a given*

visual field in the contralateral thalamus and cortex. The right dLGN receives strong input from the left eye in rodents, whose eyes have reduced visual field overlap compared to binocular species such as primates. Retinal projections segregate into eye-specific layers (or zones) in adult dLGN. These then project to the primary visual cortex (VISp). Locations of neurons with binocular RFs are marked with red and blue stripes outside the ipsilateral patch. Mouse primary visual cortex is predominantly monocular, with a smaller binocular field. Adapted from (Hooks and Chen, 2020).

Hierarchical organization of higher visual areas

The study of how the visual system is organized downstream to V1 has been mainly pursued in primates. Through anatomical and functional characterization, V1 is placed as the first cortical stage in a series of higher visual areas (HVA) that form a hierarchically organized system (Felleman and Van Essen, 1991). Areas higher up in the hierarchy are driven by convergent projections from earlier areas, to which they also send reciprocal connections. Each processing step combines inputs from the previous step to extract and integrate information across larger fields of view and represent more complex visual features. This hierarchical organization is reflected functionally with increasing size of receptive fields and complexity of preferred visual features (Hubel and Wiesel, 1965, 1979), as well as anatomically where feedforward projections terminate mainly in L4 and feedback projections avoid L4 and hierarchical ordering can be derived from the relative ratio of feedback and feedforward projections between areas (Felleman and Van Essen, 1991; Markov et al., 2014)). Visual information arriving in primary visual cortex is sent in parallel to diverging HVAs, and individual HVAs are thought to specialize in processing certain aspects of the visual inputs (Livingstone and Hubel, 1987). Furthermore, at a higher organizational level, visual areas more ventrally located in the primate visual system are more involved in object recognition (*what* pathway), while dorsally positioned areas are more action-oriented (*where* pathway), the so-called dorsal and ventral stream (Mishkin et al., 1983).

Anatomical studies have reported that mice, similarly to primates and felines, possess many higher visual areas. In fact, over 10 different areas in mouse cortex have been



reported to respond to visual input and contain an organized retinotopic map of visual space (Fig. 1.2B) (Wang and Burkhalter, 2007). Similar to felines and primates, these areas are hierarchically organized, with HVAs higher in the hierarchy responding later to visual inputs and having larger receptive fields (Burkhalter, 2016; Glickfeld and Olsen, 2017; Harris et al., 2019; Siegle et al., 2021). While first-order thalamus (dLGN) mainly provides input to V1 (and only minimally to HVAs), HVAs receive their main inputs from V1 and from higher-order thalamus (Lateroposterior area, or LP, in mice; pulvinar in primates (Baldwin et al., 2017)), which is connected to and drives activity in HVAs (Bennett et al., 2019; Blot et al., 2021; Oh et al., 2014; Roth et al., 2016; Tohmi et al., 2014). Based on their connectivity patterns, these areas are also putatively organized in ventral and dorsal processing pathways in mice (Wang et al., 2011, 2012).

Figure 1.2: Anatomical and functional organization of mouse visual cortex. ***(A)*** *Schematic overview of the dorsal part of the mouse cortex.* ***(B)*** *Triple injections into V1 and anterograde tracing of intracortical connections reveal topographically organized projection targets of V1. Area abbreviations: A: anterior, AL: anterolateral, AM: anteromedial, LI: laterointermediate, LM: lateromedial, P: posterior, PM: posteromedial, POR: postrhinal, RL: rostrolateral, S1: primary somatosensory. Panel B has been reproduced from (Wang and Burkhalter 2007).* ***(C)*** *Schematic organization with V1 surrounded by a set of HVAs. The areas shown in (B) and (C) correspond to the red dotted region in (A).* ***(D)*** *Distributions of the preferred speed of visual input in the receptive field of neurons in PM, V1, and AL. Neurons in PM neurons prefer slow-moving stimuli whereas neurons in AL prefer fast-moving input.* ***(E)*** *Bar plots of median peak speed across areas. Panel D and E were adapted from Andermann et al. 2011.*

Analogously to the primate system, in which HVAs selectively respond to motion or color, the tuning properties of mouse HVAs are area-specific and vary in temporal and spatial frequency preferences (Andermann et al., 2011; Marshel et al., 2011). In comparison to the primate visual system, however, these HVAs appear less functionally specialized and are much smaller in total surface area compared to V1 (Glickfeld and Olsen, 2017). To which extent individual areas are specialized and underlie unique cognitive processes or computations is a longstanding debate (see Chapter 5 and General Discussion).

Given the sophisticated methods available, the mouse visual system has emerged as a powerful model system to study the structural and functional organization of hierarchical visual processing. For example, it has been shown that V1 contains neurons with a wide range of tuning properties, but that the projections from V1 are specifically tuned to the preferred properties of the postsynaptic target HVAs in which they terminate and thereby contribute to areal specialization (Berezovskii et al., 2011; Glickfeld et al., 2013a). Feedback projections from these HVAs to lower areas are as abundant as feedforward projections and important questions as to what the function of these feedback projections is, remain unanswered.

Frontoparietal decision making

After initial visual processing has taken place, we may ask: how is this information integrated with the current context and subsequently used to inform decision making? Recent anatomical tracer studies by the Allen Institute have shown that a series of cortical midline structures form an important pathway for visual as well as auditory information to inform more frontally located motor systems to instruct goal-directed movements in mice (Oh et al., 2014; Zingg et al., 2014). This pathway comprises posterior parietal cortex (PPC), retrosplenial cortex, and more frontally, anterior cingulate and supplementary motor cortex (Figure 1.3). Neural responses along this pathway are increasingly multimodal, choice-related, and context-dependent; in other words, they reflect sensory features less and less, but rather correlate with how the information bears on the task

objective and the decision and action the animal is taking. These areas are reciprocally connected with higher-order thalamic nuclei forming an important cortico-thalamo-cortical network (Sherman, 2016). It has been proposed that in this network the role of the thalamus is in orchestrating transcortical communication (Halassa and Kastner, 2017; Schmitt et al., 2017). Through these pathways, different motor control regions might be recruited, depending on the required movement. For example, during behavioral tasks where licking is involved - to report choice and consume reward – different parts of the mouse motor cortex termed *anterolateral motor cortex* and *tongue-jaw motor cortex* are crucial for behavioral performance (Allen et al., 2017b; Guo et al., 2014a; Li et al., 2015; Mayrhofer et al., 2019; Svoboda and Li, 2018).

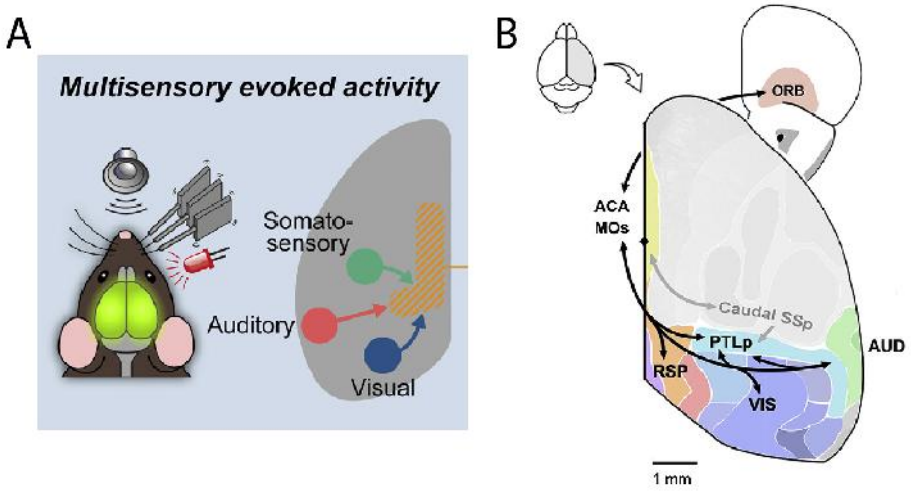


Figure 1.3: Sensory information flows converge in posterior parietal and medial prefrontal cortex to support decision making. (A) *Visual, auditory, and somatosensory inputs are relayed to medial association areas such as posterior parietal cortex and medial prefrontal cortex. The schematic was constructed based on multisensory evoked cortical network activity. Adapted from* (Kuroki et al., 2018). ***(B)*** *Sensory information is routed between sensory areas (VIS, AUD) and higher-order association areas along the medial bank of the neocortex, such as the retrosplenial (RSP), posterior parietal cortex (PTLp), anterior cingulate (ACA) and supplementary motor cortex (MOs), to areas such as orbitofrontal cortex (ORB). This medial subnetwork was identified through tracer injections placed across the entire mouse neocortex. SSP: primary somatosensory cortex. Adapted from* (Zingg et al., 2014).

Particularly the PPC is of interest to the study of flexible visual (and auditory) processing. In primates, the posterior part of the parietal cortex overlaps with several anatomically distinct areas (Hyvärinen, 1982). Compared to primates, the mouse posterior parietal cortex is strongly reduced in relative volume (as are other associative cortices). Due to its smaller size and less established anatomical work, there is a fair degree of variability in what exactly different authors refer to as constituting PPC (Hovde et al., 2019; Lyamzin and Benucci, 2019; Wilber et al., 2015). Roughly, PPC partly overlaps with earlier mentioned HVAs and is defined as including higher-order visual area A, AM, and extending to the area anterior to AM (Figure 1.2C). Neurons in rodent PPC not only respond to visual inputs but also auditory and somatosensory stimuli (Mohan et al., 2018; Nikbakht et al., 2018; Olcese et al., 2013; Raposo et al., 2014; Wallace et al., 2004). Furthermore, compared

to primary sensory areas, neurons in the parietal cortex reflect the decision animals take, rather than sensory information (Driscoll et al., 2017; Goard et al., 2016; Harvey et al., 2012; Hwang et al., 2017; Krumin et al., 2018; Nikbakht et al., 2018; Pho et al., 2018; Runyan et al., 2017; Zhong et al., 2019). Finally, PPC is reciprocally connected with primary cortices, but also strongly with supplementary motor areas (Oh et al., 2014; Zhang et al., 2016; Zingg et al., 2014). PPC is thus ideally located to integrate information from the different senses, filter relevant information, and steer action based on reward and current context.

However, despite these task-related neural correlates and its position at the interface between visual and auditory processing and premotor cortices, it is less clear under which conditions PPC causally contributes to perceptual decision making. Several studies find PPC necessary for making decisions based on visual stimuli (Driscoll et al., 2017; Goard et al., 2016; Harvey et al., 2012; Licata et al., 2017), but not for tactile (Guo et al., 2014a) or auditory information (Erlich et al., 2015; Licata et al., 2017), or find that the subdivision of PPC depends on the task modality (Gallero-Salas et al., 2021). Other studies have suggested that PPC is causally involved in both visual and auditory tasks, but only when additional cognitive processes beyond simple sensorimotor associations are required (Akrami et al., 2018; Funamizu et al., 2016; Harvey et al., 2012; Hwang et al., 2017; Licata et al., 2017; Song et al., 2017; Zhong et al., 2019). Therefore, to what extent perceptual decision making is supported by PPC remains a matter of debate and is the subject of investigation in Chapter 5.

This short anatomical and organizational overview paints an almost static picture of audiovisual sensory processing based on the circuitry-based flow of information. According to what has thus far been described, perception, cognition, and action proceed as a fixed cascade of feedforward processing steps and that - by previously made associations - invariably leads to similar neural and behavioral outcomes. However, this couldn't be less true and will be the subject of the next sections.

Brain state

A single sensory stimulus can lead to a very different neural as well as behavioral response from one moment to the next. When we are deeply asleep, a salient stimulus such as an alarm clock can elicit a behavioral response. Weak stimuli, on the other hand, are not noticed, although they would normally be perceived during wakefulness. Also during deep surgical anesthesia, stimuli are still processed by multiple neocortical areas (Alkire et al., 2008; Koch et al., 2016; Mohajerani et al., 2011; Supp et al., 2011), yet this form of evoked activity is not consciously perceived (Sanders et al., 2012). The awake state is similarly puzzling. We are able to process even the fine details of a visual scene easily, but sometimes we surprisingly fail to detect highly salient objects (Simons and Chabris, 1999) or major changes in the visual scene if these coincide with some visual disruption such as an eye movement or a brief obscuration of the image (Grimes, 1996; Simons and Levin, 1997).

What are the mechanisms underlying such a high variability in the way the same sensory stimulus is processed, and how are they relevant for understanding perception? A key factor to consider is the interaction between brain state and signals originating from sensory transducers. While the latter have been extensively investigated, the nature of brain states and how they influence sensory processing has received surprisingly limited attention. How is sensory processing shaped by the characteristics of brain states?

#Parts of the section 'Brain State' were previously published in Olcese, Oude Lohuis, Pennartz, 2018

Synchronized and desynchronized brain states

Brain states have been classically distinguished based on the striking differences in aggregate electrical activity (for example using electroencephalography, EEG). Wakefulness is classically characterized by the presence of desynchronized neural activity (McGinley et al., 2015a), while NREM sleep (as well as forms of non-dissociative anesthesia, i.e. anesthetics involving general loss of consciousness) is instead characterized by massive neural synchrony, also referred to as slow wave sleep (SWS). It is not difficult to imagine that sensory inputs will have different futures depending on whether they arrive in a system of collective firing, alternating with collective silence (synchronized state), or when there is sparse balanced activity (desynchronized state).

A host of studies has characterized sensory responses across synchronized and desynchronized states. Consistently, single neurons in primary sensory areas show remarkably similar tuning to stimulus features across behavioral states, e.g. neurons in visual cortex rarely shift their preferred orientation (Durand et al., 2016; Ecker et al., 2014; Goard and Dan, 2009; Goltstein et al., 2015; Niell and Stryker, 2010; Nir et al., 2015). In general, responses seem to have slightly faster dynamics in desynchronized brain states (i.e. shorter onset latency and faster transients) (Durand et al., 2016; Haider et al., 2013; Hasenstaub et al., 2007; Wang et al., 2014; Wörgötter et al., 1998), but see (Pachitariu et al., 2015)). Even though this suggests that major feedforward pathways are to a large extent functioning similarly throughout brain states, response dynamics can vary considerably when considering the immediate local state that sensory inputs face when reaching the cortex.

Early work in the anesthetized cat visual cortex showed that identical stimuli can elicit different neuronal responses depending on whether they arrive during UP or DOWN states (Arieli et al., 1996; Azouz and Gray, 1999; Haider et al., 2007), but see (Haider et al., 2013)). Also in the awake rodent somatosensory cortex responses to single whisker deflections are larger when occurring during DOWN states of synchronized activity, as opposed to UP states (Crochet and Petersen, 2006; Hasenstaub et al., 2007; Petersen et al., 2003; Sachdev et al., 2004; Sachidhanandam et al., 2013), similar to responses of neurons in auditory cortex to isolated tones (Deweese and Zador, 2004; Sela et al., 2016). Across repeated presentations of the same stimulus this leads to increased trial-to-trial variability, compared to the desynchronized state, during which fluctuations in population activity and network excitability are instead smaller (Haider et al., 2013; Zagha et al., 2013).

State variability within wakefulness

Also within wakefulness sensory processing dynamically varies alongside fluctuations in brain state. It is important to note that wakefulness is hardly definable as a single behavioral state. When we are awake, we can either sit quietly and mind-wander or be highly involved in a myriad of different activities, from running to thinking. A key factor which varies during wakefulness is the arousal level (McGinley et al., 2015b, 2015a; Reimer et al., 2014), as a function of which different levels of cholinergic and noradrenergic activity (amongst other neuromodulators) modulate baseline neuronal activity and consequently the way in which sensory stimuli are processed.

The arousal level is primarily measured in terms of pupil diameter, with a larger pupil being associated with higher arousal (McGinley et al., 2015a; Reimer et al., 2014). High arousal

is accompanied by desynchronized cortical activity, while during low arousal neural patterns become more synchronized (McGinley et al., 2015a), to the point that slow oscillations (the hallmark of NREM sleep) occur during quiet wakefulness (McGinley et al., 2015b; Petersen et al., 2003; Sachidhanandam et al., 2013). Locomotion usually corresponds to a state of high arousal (measured in terms of pupil size), although the two are not unequivocally linked, at least as far as visual processing is concerned (Vinck et al., 2015a). During locomotion, visual responses are, similar to periods with large pupil diameter, enhanced (Dadarlat and Stryker, 2017; Kaneko et al., 2017; Niell and Stryker, 2010), yet via a distinct mechanism. While pupil-related arousal suppresses spontaneous firing activity in V1 (thus promoting the emergence of stimulus-evoked responses), locomotion increases stimulus-evoked activity (Bennett et al., 2013; Dadarlat and Stryker, 2017; Fu et al., 2014; Polack et al., 2013; Vinck et al., 2015a). Through both of these mechanisms the consistency and signal strength in the coding of stimuli appears to be enhanced – but see (Shimaoka et al., 2018). Across identical stimulus presentations the variability in response is reduced when the arousal level is high (McGinley et al., 2015b; Polack et al., 2013; Schneider et al., 2014; Schölvinck et al., 2015) with again a key contributor being reduced pre-stimulus variability (Bennett et al., 2013; Zagha et al., 2013). This, in combination with higher membrane conductance in the desynchronized state (Wang et al., 2014) or during locomotion (Bennett et al., 2013), can lead to an increased signal-to-noise ratio (response versus baseline variance) (Bennett et al., 2013; Pachitariu et al., 2015; Vinck et al., 2015a). These state-dependent alterations in single neuron coding correlate with increased performance in sensory detection tasks (Bennett et al., 2013; McGinley et al., 2015b; Pinto et al., 2013), but see (Sachidhanandam et al., 2013).

High noise correlations during synchronized activity

As we discussed, the information-signaling capacities of single neurons increase during desynchronized activity. However, whether a single neuron responds to a stimulus – which still occurs during deep anesthesia – is not informative in explaining why some stimuli come to be perceived or not. As stimuli are likely coded by patterns of ensemble activity (Pennartz, 2015; Pouget et al., 2000), it is essential to understand at a population level how the representation of information varies across states. We will here discuss the impact of correlated variability between neurons for accurate perception.

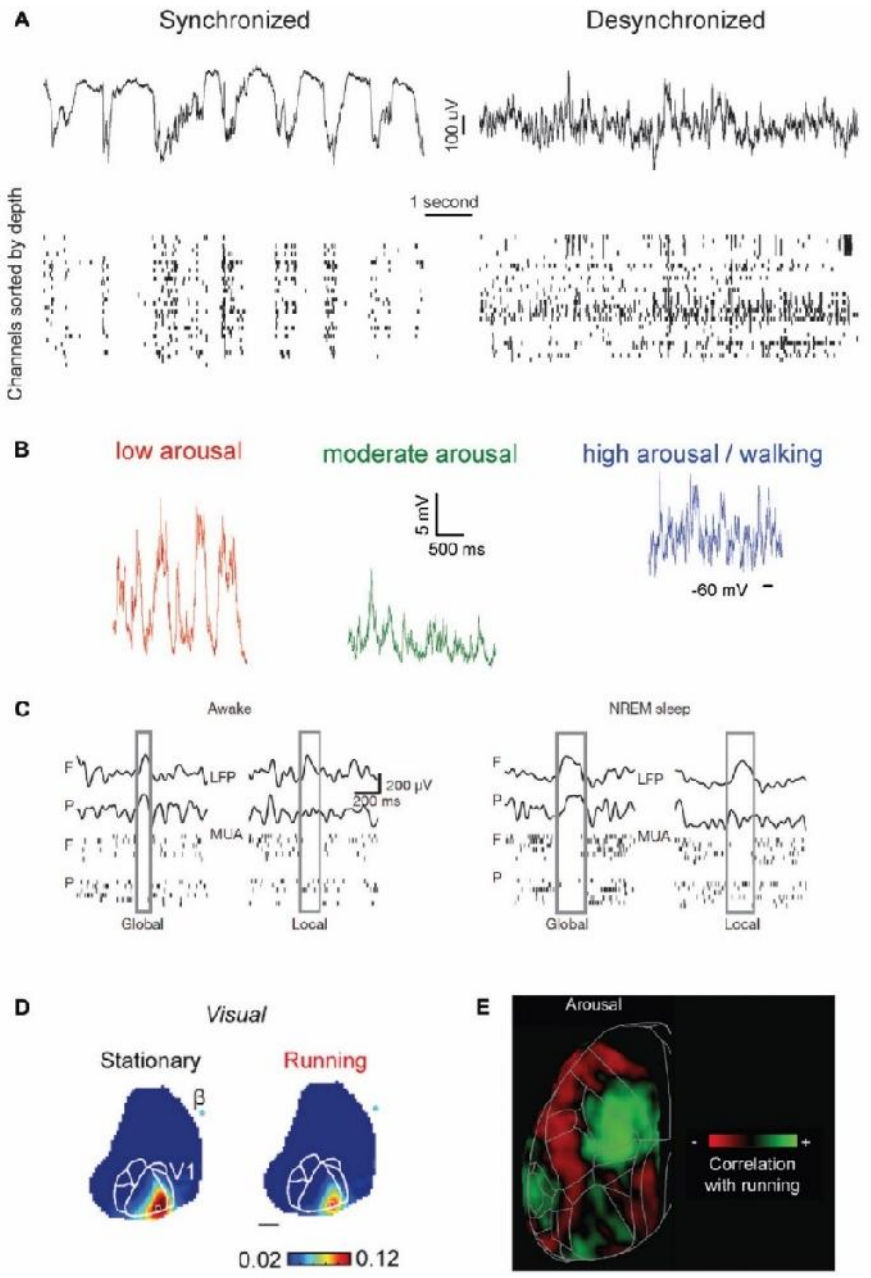


Figure 1.4: An updated definition of brain states. (A) *Brain states have traditionally been distinguished based on the characteristic of population-level oscillatory dynamics present within a behaviorally defined homogeneous period. Left, top: local field potential (LFP) trace present in mouse visual cortex during isoflurane anesthesia. LFP activity is characterized by oscillatory dynamics with strong power in slow frequencies (0.5–4 Hz). Left, bottom: at the neuronal level, isoflurane anesthesia determines the alternation of periods of spiking and silence which are synchronous throughout cortical areas, and in phase with the co-occurring LFP oscillations. Right:*

same as left panel, but during wakefulness. Note the disappearance of slow frequency oscillations at the LFP level and the overall loss of synchrony for neuronal activity. ***(B)*** *Recent studies have shown that, within wakefulness, much more variability is present than previously thought. In periods characterized by a low arousal level (left) neuronal activity (here shown in terms of intracellular membrane potential traces) displays patterns indistinguishable from those present during non-REM sleep or anesthesia. As the arousal level increases, or if a period of activity—locomotion—occurs (center, right) activity becomes more and more desynchronized. The tick mark indicates -60 mV as a reference for the voltage traces.* ***(C)*** *During Non-REM sleep and forms of non-dissociative (e.g., isoflurane) anesthesia UP and DOWN states normally occur globally, i.e., involving the whole thalamocortical system. However, UP/DOWN states can also be local events, involving one set of cortical regions. This mostly occurs during wakefulness when homeostatic sleep pressure is high (left), and during non-REM sleep when, conversely, sleep pressure is slow. Each plot shows an example of global or local DOWN state occurring during wakefulness (left) or sleep (right), and measured in terms of both LFP recordings or neuronal multi-unit activity (MUA). F: frontal derivation. P: parietal derivation.* ***(D)*** *Sensory processing varies within wakefulness. Visual evoked potentials (here shown as measured by voltage-sensitive fluorescent proteins, scale bar indicates fluorescence relative to baseline) are generally weaker when a mouse is running compared to when it is stationary. The same is observed for audition and somatosensation. β: Bregma.* ***(E)*** *Arousal differentially affects cortical areas in the mouse. Green indicates cortical areas where the signal measured via voltage-sensitive fluorescent protein is positively correlated with arousal (locomotion). Red shows areas where a negative correlation is present. D, E: Adapted from Olcese et al. 2018. Panel (A) was produced by M.N. Oude Lohuis (UvA), (B) adapted from McGinley et al. (2015b), (C) adapted from Vyazovskiy et al. (2011), (D) and (E) adapted from Shimaoka et al. (2018).*

Upon rhythmic oscillations in population firing, spiking activity is positively correlated between nearby neuronal pairs, i.e. neurons increase and decrease their firing together (Lampl et al., 1999). Oppositely, during a desynchronized state, neurons do not show this large synchronous modulation of firing rate and fire more or less independently (Renart et al., 2010). If these synchronous fluctuations persist in responses to stimuli, i.e. if neurons show correlated variability across repeated presentations of the same stimulus, these pairwise correlations are called noise correlations (Abbott and Dayan, 1999; Averbeck et al., 2006; Cohen and Kohn, 2011), in contrast to signal correlations (pairwise correlation in the average response across stimuli).

Noise correlations are generally high during anesthesia or otherwise synchronized states (Ecker et al., 2014; Renart et al., 2010) due to these common fluctuations in activity, and primarily arise as a result of neurons transiently ceasing firing together (Mochol et al., 2015). However, noise correlations increase even under light anesthesia and in the absence of clear UP / DOWN states (Golshani et al., 2009; Goltstein et al., 2015; Greenberg et al., 2008), and also when focusing only on UP states (Renart et al., 2010). Noise correlations can also vary within wakefulness (Gentet et al., 2010; Poulet and Petersen, 2008). Specifically, noise correlations decrease upon locomotion (Erisken et al., 2014; Vinck et al., 2015a) or arousal induced by an air-puff (Vinck et al., 2015a).

Desynchronization improves population coding

Traditionally, noise correlations have been argued to impair sensory processing (Shadlen and Newsome, 1998; Zohary et al., 1994), but recent theoretical work has shown that only certain noise correlations (termed differential correlations) are information limiting, namely those fluctuations that are identical to those generated by stimulus variations

(Kohn et al., 2016; Moreno-Bote et al., 2014). Additionally, our previous work suggests that pairwise noise correlations are not inherently detrimental (Montijn et al., 2014), especially when considering them within the context of larger populations (Montijn et al., 2016). This is because multidimensional noise correlations (covariability between neuronal triplets etc.) are more likely to be orthogonal to the dimension coding stimulus identity, limiting the effect on population readout in downstream areas (Montijn et al., 2016).

Nonetheless, high noise correlations during synchronized activity do impair population coding. In macaque visual cortex a locally synchronized state just before stimulus onset impairs population readout of stimulus identity from a recorded ensemble and impairs behavioral performance as well (Beaman et al., 2017), and the amount of decorrelation upon visual input is correlated with detection of figure-ground stimuli (van der Togt et al., 2006). Experimentally promoting cortical desynchronization by electrical stimulation of the nucleus basalis (containing mostly cholinergic neurons) in rats improves population decoding of natural scenes (Goard and Dan, 2009), and performance in a visual discrimination task (Pinto et al., 2013). This suggests that increased desynchronization facilitates independent coding of stimulus features and that such heterogeneous neuronal responses contribute to accurate perception. Indeed, the amount of relative contrast in activity between neurons, indexed as population response heterogeneity, correlates with hits versus misses in a visual detection task (Montijn et al., 2015). This study showed that a highly heterogeneous activity pattern even preceded detected stimuli, whereas heterogeneity was lower during the pre-stimulus baseline of miss trials. This suggests that task epochs are characterized by different levels of heterogeneity in neural activity and that these fluctuations might impact how incoming stimuli are subsequently processed.

While the improved population coding during desynchronized activity might contribute to accurate perception in sensory detection tasks, it is still present within non-conscious animals, such as when comparing desynchronized with synchronized epochs during urethane anesthesia (Goard and Dan, 2009; Pachitariu et al., 2015). Therefore, whether desynchronization *per se* contributes to perception remains to be addressed. This likely depends as well on whether this decorrelated and more independent coding of sensory features also actually benefits downstream neurons. Chapter 3 addresses how decorrelation of V1 population activity might contribute to improved readout in combination with perturbations of V1 activity with variable timing relative to population decorrelation.

Brain state, and in particular the degree of synchronized fluctuations in activity shape the feedforward processing in important ways. However, it is particularly contextual modulations that are dependent on long-range and recurrent interactions with V1 are affected, or even lost, during sleep or under anesthesia (Adesnik et al., 2012; Haider et al., 2013; Keller et al., 2020a; Vaiceliunaite et al., 2013). Anesthesia can therefore be used as a manipulation to study state-dependent brain changes in the context of loss of consciousness (Alkire et al., 2008; Schröter et al., 2012; Untergehrer et al., 2014). In Chapter 2 we also investigate how the function of feedback from higher-order visual areas to V1 might be different under states of anesthesia versus wakefulness.

Task setting

The previous section explicitly addressed how brain state and arousal can radically alter sensory processing. However, this is not necessarily a homogeneous process for all potential stimuli. Some encountered stimuli will have value for an organism. For example, you might wake up more easily from hearing your own name, but not that of others. The ability to modify sensory processing based on which stimuli are predictive of reward or danger is central to adaptive brains. How behaviorally relevant stimuli shape sensory processing can be broadly divided in two ways: reward-dependent structural network reconfigurations and rapid attentional and engagement effects.

Reward-dependent learning in sensory systems

Sensory systems do not come prewired, ready to analyze the world, but are wired through repeated interactive experience with the world. Critical periods of enhanced plasticity are specifically instructive to the appropriate wiring of sensory systems (Hooks and Chen, 2020; Levelt and Hübener, 2012). However, even during adulthood positive and negative outcomes are still able to adapt sensory processing to facilitate the analysis of behaviorally relevant features in the world. Selective stimulus-outcome learning can alter the selectivity of neuronal representations at both a single neuron and population level all the way down to early sensory cortices (Bao et al., 2001; Fritz et al., 2003; Ghose et al., 2002; Goltstein et al., 2013, 2018; Henschke et al., 2020; Poort et al., 2015; Schoups et al., 2001). At the single neuron level, the (bottom-up) saliency of a conditioned stimulus might be enhanced by a selective increase of response amplitude (Blake et al., 2006; Fritz et al., 2003; Goltstein et al., 2013; Poort et al., 2015), by the recruitment of additional neurons that become tuned to the conditioned stimulus (Weinberger et al., 1993), or alterations in the tuning curves (Goltstein et al., 2013; Schoups et al., 2001; Yang and Maunsell, 2004). At the population level, reward-dependent learning can increase the sparseness of population responses to the rewarded stimulus (Gdalyahu et al., 2012; Ghose et al., 2002) or reshape the correlation structure of ensemble activity (Averbeck et al., 2006; Ghose et al., 2002; Jeanne et al., 2013; Montijn et al., 2015). Even in the absence of an active behavioral task, alterations of stimulus processing in sensory cortex can result from simple repeated reward pairing (Henschke et al., 2020; Seitz and Watanabe, 2009). Non-visual feedback or neuromodulatory inputs have been suggested to form the substrate for learning experiences to drive plasticity and alter V1 local circuitry (Bao et al., 2001; Chubykin et al., 2013; Gilbert and Li, 2013). Perceptual learning might also be associated with reconfigurations of long-range interactions. In the somatosensory system, texture discrimination was associated with strengthened connectivity specifically from S1 to S2, highlighting that behavioral relevance might affect how information is routed from primary to downstream areas (Chen et al., 2016). The behavioral relevance of specific sensory stimuli thus leads to large-scale modifications in sensory processing through experience.

Attention and task engagement

However, stimuli are generally only relevant under certain conditions, rarely under all circumstances. Rather than long-term changes across learning, mechanisms must be in play to transiently modulate processing depending on moment-to-moment relevance of sensory inputs. This study of how the brain achieves dynamic control over which sensory features require preferential processing over others (i.e. attention) is a research field on its own. A specific focus of this thesis is the question of how sensory processing is different for

a stimulus that is directly predictive of an available reward, versus the same input presented outside of this rewarding context.

In primary sensory cortices, it is often found that a large fraction of neurons responds similarly to sensory inputs during task engagement, as well as to stimuli presented outside the task setting. For example, mouse V1 neurons show similar sensory-evoked activity to visual stimuli presented after task performance when the lick spout is removed (Pho et al., 2018; Steinmetz et al., 2019). Primate S1 neurons also faithfully track tactile inputs also outside the task (Hernández et al., 2010). However, the fraction of purely sensory-driven neurons decreases along the sensorimotor hierarchy. For example in the primate somatosensory and visual system, the fraction of neurons that cease to respond to the stimuli outside the task and that are instead modulated by task engagement and the perceptual decision process increases along the hierarchy (Romo and Rossi-Pool, 2020; Rossi-Pool et al., 2021; Siegel et al., 2015). Likewise, in ferrets, stimuli that are rewarded (and therefore more salient and attended to) are enhanced in terms of response amplitude and evoke enhanced responses, especially in higher-order areas enabling them to better propagate to frontal areas (Fritz et al., 2010; Yin et al., 2020). Similarly, mouse parietal cortex only showed task-related signals in trials in which the visual stimulus could be used to obtain reward, but not during blocks when the lick spout was removed (Pho et al., 2018). Overall, the effect of task engagement thus increases along the sensorimotor hierarchy. It must be noted that also in early sensory areas, for example in V1, signals are reported that index engagement (Steinmetz et al., 2019) or whether the context is associated with visual rewards (Hajnal et al., 2021). Also in subcortical structures task engagement can amplify neuronal responses to rewarded target stimuli, for example in inferior colliculus of primates (Shaheen et al., 2021) or mice (De Franceschi and Barkat, 2021), but these effects are often much smaller compared to higher-order areas.

Task engagement thus has a striking effect on how sensory inputs propagate across the brain. The modulation of sensory processing by task context might thus be for an important part achieved through modification of the communication of lower-order sensory areas with higher-order areas and affecting the routing between areas. Large open questions remain on how selective propagation of task-relevant information is controlled and implemented. Again, top-down inputs, as well as neuromodulatory influences, could provide the substrate for directing attentional resources to the selective processing of relevant stimuli (Desimone and Duncan, 1995; Maunsell, 2015; McAdams and Reid, 2005; Rombouts et al., 2015; Zhang et al., 2014), and reshape how areas communicate to allow sensory information to instruct appropriate actions (Fries, 2005; Kaufman et al., 2014; Kohn et al., 2020).

In sum, the processing of relevant stimuli can be modified through long-term structural changes as well as more dynamic mechanisms. However, identifying whether modulations of sensory-evoked activity are due to long-term experience, task engagement, or more brain state-related effects is not trivial. The effect of task engagement is often studied in animals that are highly trained for a given perceptual task. Furthermore, rewards by themselves evoke signals in primary sensory areas, either directly or through reward anticipation (Lacefield et al., 2019; Shuler and Bear, 2006). Second, to complicate matter further, rewards, as well as movements, cause increased arousal, which are both associated with pupil dilation, and in the visual cortex with increases in average firing rate and

decorrelation of activity (McGinley et al., 2015b; Vinck et al., 2015a; Zagha et al., 2013). All of these factors are associated with substantial alterations in visual processing. It is therefore challenging to disentangle the distinct effects of rewards, arousal, and movement in the average rodent perceptual task (Maunsell, 2004). This is because animals are trained to restrict movements to report a stimulus when a reward is expected and arousal levels increase; therefore, in perfectly trained animals accurate perception, reward expectancy as well as reward delivery, consummatory movements and arousal are tightly correlated and hard to isolate. Disentangling the contribution of all these variables, requires controlled manipulation of task engagement, the timing of reward delivery, as well as careful monitoring of arousal (pupil size) and body movements.

Crossmodal influences on visual cortex

Finally, vision does not operate in isolation, nor does any other sense. Our brains, as well as nervous systems across the animal kingdom, are continuously sampling the world in an inherently multisensory manner (Flanigan et al., 2021; Ghosh et al., 2017; Solvi et al., 2020; Stein and Stanford, 2008). Our perception of flavor, for example, involves the integration of gustatory and olfactory cues (Maier et al., 2015; Spence, 2015) and we use visual lip movements to understand better what someone is saying (Ross et al., 2007; Schroeder et al., 2008; Sumby and Pollack, 1954). Each sense provides unique information and properly integrating, segregating, selecting, and ignoring multisensory cues improves our understanding of the world.

Forms of multisensory processing

Multisensory processing research has primarily focused on *cue integration,* the process by which the integration of external cues from multiple modalities can provide a more reliable estimate of an object or event as compared to unimodal cues (Fetsch et al., 2013; Gleiss and Kayser, 2012; Lippert et al., 2007; Meijer et al., 2018, 2020; Meredith and Stein, 1986). For example, we might judge the shape of an object by combining our estimates based on touching and visually observing the object itself (Ernst and Banks, 2002). However, cue integration is only one of the many types of multisensory processing (Meijer et al., 2019), and additional processes determine the interplay of the senses such as sensory prioritization (Lakatos et al., 2009; Lee et al., 2016; Wimmer et al., 2015) and resolving multisensory conflicts (Song et al., 2017). Even more importantly, while the field has mostly focused on the behavioral and neural correlates of cue integration, most multisensory events in daily life do not share the same underlying cause and should be segregated rather than integrated (Körding et al., 2007; Mihalik and Noppeney, 2020; Noppeney, 2021; Rohe and Noppeney, 2015; Shams and Beierholm, 2010). For example, the sound of a honk of a car on the street bears little on the processing of all the other objects in the scene. In a slightly different vein, it can be argued that it is equally important to segregate the modality of the stimuli to account for the qualitatively varied nature of conscious experience (Pennartz 2009; Pennartz, 2015). Both visual and auditory stimuli evoke patterns of distributed spiking activity, but are associated with a different character of experience. According to the latter framework, a form of crossmodal interaction and calibration is necessary that preserves the informational content in both domains, but differentiates visually from auditory induced spiking activity to account for the fact that both modalities have a segregated, distinct qualitative experience. This hypothesis is central to the design of the behavioral task presented in Chapters 3, 4, and 5. Specifically,

we aimed at investigating whether auditory changes would lead to activity changes in the visual system without the purpose to enhance visual feature representation, but rather to engage crossmodal audiovisual interactions, hypothesized to shape the nature of the perceived experience (Pennartz, 2009).

To implement these multisensory processes flexibly, we are coming to the understanding that interactions between the senses occur at most, if not all, levels of the sensory processing hierarchy and in different forms (A. E. Allen et al., 2017; Bieler et al., 2018; Ghazanfar and Schroeder, 2006). Thus, contrary to the initial sketch of perceptual decision making introduced earlier, multisensory processes are not restricted to dedicated higher-order areas such as posterior parietal and frontal cortices, but occur in early as well as late stages of sensory processing. Therefore, the neural correlates of multisensory processing are expected to depend on the level under study and on the behavioral relevance and goals (top-down factors) and not only on bottom-up factors such as spatial and temporal coincidence (Choi et al., 2018; De Meo et al., 2015; Meijer et al., 2019; van Atteveldt et al., 2014).

Audiovisual interactions

Although multisensory interactions occur between the various senses, most research has focused on audiovisual interactions. Vision and audition are important senses that inform animals about distal events with high spatiotemporal resolution. Vision and audition capture fundamentally different physical signals (sound waves versus photons) and it is to no surprise that they convey different aspects of the reality around us. The most relevant information in sounds is in the temporal domain and extracting this information requires highly temporally specific processing. For example, inter-aural time delays which are used to estimate sound source location depend on microsecond differences between sound wave impinging on the ears. Oppositely, visual inputs are rich in information in the spatial domain and the visual system is thus designed to care about local spatial contrasts, rather than fast temporal changes in inputs. The auditory system, therefore, has a high temporal resolution and is accurate in estimating timing and onsets – i.e. *when something happens* - while the visual system is optimized to provide spatially precise estimates – i.e. *where something happens*. This contrast is relative, as humans can also easily perform localization using audition and temporal estimation using vision. Nonetheless, this relative specialization is evident in multisensory illusions. The sound-induced double flash illusion depends on a *temporal* estimate and sounds can make subjects falsely perceive flashes (Shams et al., 2000). The ventriloquist illusion, on the other hand, depends on a *spatial* estimate, and vision modulates the perceived origin of sounds (Howard and Templeton, 1966).

Vision and audition thus convey different signals about the world and multisensory interactions are diverse in nature. With this in mind, it is particularly interesting to ask how auditory signals could inform visual processing and at which stage. Indications of audiovisual interactions have been found in the superior colliculus, higher-order associative areas, as well as primary sensory areas. Reported forms of interaction include neurons whose response is modulated by simultaneous audiovisual presentation and neurons that respond to both modalities and show (non)linear responses to simultaneous audiovisual inputs. In the context of this thesis, modulation of early visual processing by sounds is particularly relevant. Sounds can modulate visual response properties of V1

neurons such as orientation and contrast tuning (Ibrahim et al., 2016; Meijer et al., 2017). Neurons in V1 even show direct responses to sounds (Meijer et al., 2017) and selectivity to auditory frequency (Knöpfel et al., 2019). These modulations have been shown to depend on direct projections from auditory cortex (Ibrahim et al., 2016; Iurilli et al., 2012), with auditory cortical neurons sensitive to sharp sounds onsets being more likely to project to V1 than neurons sensitive to slowly evolving features (Deneux et al., 2019). This suggests that sounds activate visual cortex through fast horizontal connectivity via auditory cortex. These auditory signals in V1 have been mostly studied in passive animals and in the framework of cue integration. Major questions remain about the functional organization of these auditory inputs, how auditory signals are integrated with visual inputs, and how this supports or interacts with vision. Furthermore, as mentioned above, the nature and function of different forms of multisensory processing significantly depends on the behavioral context in which they occur. Furthermore, sounds can lead to motor activity, which, through direct modulation of V1 activity, can obfuscate the interpretation of sound-evoked activity (Bimbard et al., 2021). It is therefore unclear which aspects of the sound-evoked activity reflect sensory features, or rather arousal or behavioral components, all questions that are addressed in Chapter 4.

Methodological considerations

To investigate how contextual modulations affect visual processing, one needs to carefully consider how to study the unique modulatory effect of each of the considered factors on visual perception in a way that allows to narrow down the functional motifs, but also the methods to causally test hypotheses. To address this, we develop an integrated approach including behavioral manipulations, electrophysiological recordings and causal manipulation of neural activity.

Behavioral setup

Our central questions revolved around how visual processing is dependent on contextual factors. To answer this question, experiments in head-fixed mice are ideal because we have tight control over the sensory stimuli, can train them to report and use certain stimuli to perform certain behavioral actions, and vary the task contingencies while recording and manipulating the circuitry under different brain states (anesthetized, quiescent, or behaving). Furthermore, mice offer a suitable model to study sensory processing in relation to consciousness (Storm et al., 2017).

In Chapter 2 we studied hierarchical visual processing in awake or anesthetized mice that were passively presented with stimuli. In other words, they were not required to report or act upon the stimuli. For the experiment in chapters 3, 4, and 5, on the other hand, we developed a multisensory change detection task. To study the modulation of visual processing we kept the stimuli unchanged and varied, across cohorts of mice, whether audition and vision were task-relevant or not. We compared 3 cohorts of mice: naive animals, animals trained to detect visual changes but ignore auditory changes, and lastly, animals trained to detect and discriminate auditory and visual changes in the full version of the task. In the full multisensory version, mice were required to continuously and simultaneously monitor audiovisual stimuli and respond to one side for changes in auditory frequency and to the other side for changes in the orientation of a drifting grating. In other words, mice were trained to detect changes and identify the sensory modality in which the

change occurred. In trained animals, we varied task difficulty by varying the amount of change relative to the perceptual threshold. These psychophysical methods allow us to make inferences about perception and decision making by systematically relating sensory information, neural responses, and perceptual report (Carandini and Churchland, 2013; Green and Swets, 1966; Stüttgen et al., 2011).

We designed this task for various theoretical and experimental reasons. First, by using changes in a feature of the stimuli (orientation or auditory frequency) we aimed at making the task reliant on feature-processing and thus cortex-dependent (see Chapters 3 and 4). Second, by changing features of ongoing stimuli we minimized the effects that sudden and salient onsets and offsets of sounds and high contrasts have on both neural activity and arousal (Deneux et al., 2019; Iurilli et al., 2012; Yeomans and Frankland, 1995). Furthermore, in this way, we maintained constant visual luminance and contrast, which meant that observed pupil size variability indexed arousal fluctuations rather than luminance (Mathot, 2018; McGinley et al., 2015a). Third, this task goes beyond Go/No-go tasks as animals cannot simply respond to any input (Carandini and Churchland, 2013; Guo et al., 2014b; Meijer et al., 2018). Instead, the animal needs to not only detect but also identify the modality in which the change occurred, hypothesized to require long-range intracortical signaling differentiating modality-specific representations (Pennartz, 2015, 2009). Fourth, simultaneous visual and auditory detection permitted us to see how visual and auditory detection involved distinct or shared neural populations (Masset et al., 2020; Nikbakht et al., 2018; Raposo et al., 2014; Vergara et al., 2016). Lastly, Go/No-go tasks suffer from an asymmetry in motor outputs; comparing a withheld still trial (no-go) with a trial with active behavior, reward expectation, and licking (go trial) conflates choice signals with those associated with vigorous licking and arousal. Therefore having two-alternative motor actions in response to the same sensory stimuli allowed us to better dissociate choice from licking (Burgess et al., 2017; Guo et al., 2014b; Runyan et al., 2017; Steinmetz et al., 2019).

Electrophysiological recordings

This task design thus allowed us to answer our research questions by manipulating how contextual factors modulate visual processing while retaining the same visual and auditory inputs. We were interested in how horizontal intracortical connections and feedforward and feedback information flows within or outside the visual system convey contextual or task-related information. We opted for multi-area laminar probe recordings to dissect this circuitry as, first, this permitted monitoring the spiking activity of a large number of neurons across different cortical areas (Jun et al., 2017; Steinmetz et al., 2019). Second, laminar extracellular recordings provide good temporal resolution to distinguish moments of feedforward and feedback processing (Allman et al., 1985; Lamme and Roelfsema, 2000; Roelfsema et al., 1998), as well as an idea of the laminar distribution to isolate the anatomical origins of feedforward and feedback (Bastos et al., 2015; Nandy et al., 2016; Senzai et al., 2019). Furthermore, in the mouse >90% of narrow spiking neurons are parvalbumin-expressing (PV) inhibitory neurons and >90% of broad spiking neurons are excitatory neurons (Lee et al., 2010; Pfeffer et al., 2013; Rudy et al., 2011) and thus allow to classify neuron types (Niell and Stryker, 2008, 2010; Vinck et al., 2015b) (Figure 1.5).

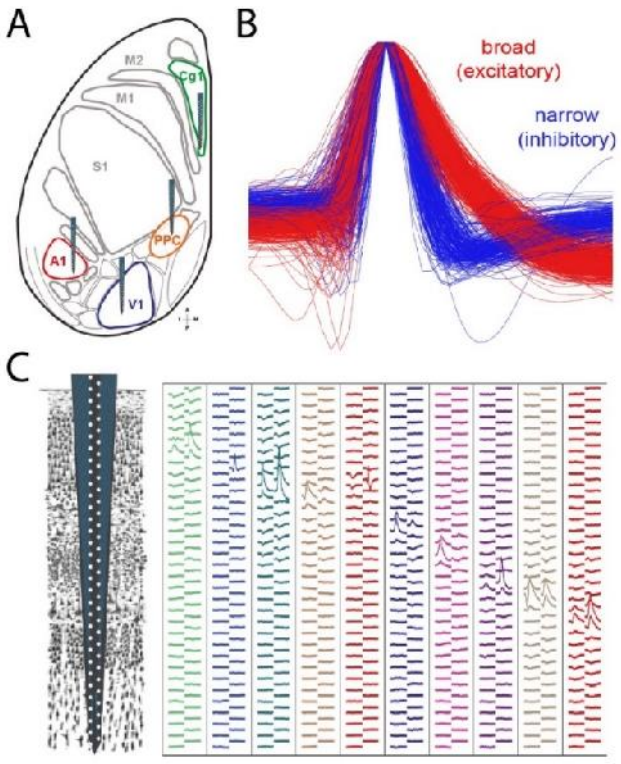


Figure produced by M.N. Oude Lohuis (UvA).

Figure 1.5: Multi-area ensemble recordings with silicon probes. (A) *Dorsal view of mouse cortex with areas of interest for neural recordings and manipulations highlighted. Data from cingulate cortex (Cg1) is not presented in this thesis. A1: primary auditory cortex; S1: primary somatosensory cortex.* ***(B)*** *Example waveforms of narrow spiking (putative fast-spiking inhibitory neurons) and broad spiking neurons (putative excitatory neurons) in V1.* ***(C)*** *Recordings were performed across the layers of mouse cortex. The left panel shows a silicon probe overlaid on a schematic of the cortical laminae. The right panel shows the voltage trace during action potentials for some example neurons (in different colors) across all electrode channels (positioned in the same zigzag configuration as the probe schematic in (A)). Neurons recorded at different cortical depths evoke voltage deflections at localized silicon probe channels.*

Causal interventions

To test the causal contribution of the neural circuits under investigation we implemented two causal intervention techniques, muscimol injection and optogenetics. Muscimol is a potent, selective agonist for the $GABA_A$ receptor and when infused in the targeted region of interest transiently and reversibly silences neuronal activity. Optogenetics involves the use of light to control neurons that have been genetically modified to express light-sensitive ion channels. We optogenetically inactivated single cortical regions by virally expressing cre-dependent ChR2 in PV-cre mice. Subsequent local photostimulation drives PV inhibitory interneurons and effectively silences activity (Madisen et al., 2012). Muscimol manipulations have poor temporal specificity and the possibility of adaptive mechanisms kicking in to compensate for the loss of function should be considered. Optogenetic manipulation, on the other hand, has far superior temporal control and was leveraged to disrupt specific temporal windows during sensory processing. However, care should also be taken that the immediate disruptive manipulations has no off-target effects (Otchy et al., 2015).

Outline of the thesis

Non-visual signals and modulatory influences are thus pervasive throughout the visual cortex and visual perception appears strongly dependent on sounds, state, and setting

(Figure 1.6). To understand why these signals are present we have to consider that, for visual input to make sense, a lot of additional contextual information needs to be considered and integrated. For example, to know whether *objects in the world* are moving or whether it is *you* that is moving, the visual system needs information about your own locomotion. To understand what we are seeing we need to compare and verify visual information with other modalities, and to act we need to know what is behaviorally relevant. The process of vision thus involves the active construction of internal models of the world using sources beyond the eyes and depends on the behavioral and perceptual needs of the animal.

Contextual influences modulate visual processing via top-down projections from higher brain areas, intra-areal horizontal connections, or neuromodulatory inputs. However, knowledge of the circuit mechanisms that integrate feedforward sensory information with contextual information is still very limited. Likewise, we do not understand how these might improve or deteriorate visual processing and visually-guided behavior. This thesis investigates how various contextual factors are encoded in visual cortex, how they interact with visual input, and how they might support visual perception. We specifically aimed to understand how visual processing and decision-making is co-determined by:

- Internal feedback: recurrent signals from higher-order areas (Chapters 2, 3, 4)
- State: brain conditions such as anesthesia and wakefulness (Chapters 2-5)
- Multisensory signals: auditory and tactile influences (Chapters 3 and 4)
- Task setting: rules, goals and context (Chapter 3, 4, 5)

First, in **Chapter 2**, we study feedback connections from higher-order to lower-order areas in the visual system. Feedback connectivity is thought to convey contextual information to facilitate lower-order processing and has been linked to attention, prediction, and awareness. It is unclear how different higher visual areas functionally modulate and shape visual response properties in V1. As the functionality of feedback projections is thought to be selectively affected by state changes (anesthesia vs. wakefulness), an additional question is therefore whether this functional organization depends on brain state.

In **Chapter 3**, we investigate whether V1 activity and its contribution to visual perception are affected by (multisensory) task setting. Several studies have shown that V1 responses are influenced by the larger visual context. We were interested in whether V1 dynamics were also affected by the presence of task demands in another modality and tested whether the role of V1 is fixed for a given stimulus, or can be better conceptualized as a recurrent node in the network. Finally, we ask if later activity patterns are still causally required for visual perceptual decision-making.

In **Chapter 4**, we address a recent controversy regarding the nature of sound-evoked responses on V1. Both auditory and motor movements have been shown to influence V1. However, it is unclear to what extent sounds cause movements that are subsequently observed in V1, nor is it clear how auditory and motor-related inputs jointly shape visual processing. In Chapter 4 we use multi-area recordings, task manipulations, pharmacological interventions and optogenetics to dissociate visual, auditory and motor processing during multisensory evoked activity in V1.

Lastly, **Chapter 5** moves beyond the primary visual cortex (V1) and focuses on the role of the Posterior Parietal Cortex (PPC). To what extent perceptual decision making is supported by PPC remains a matter of debate, with specific controversy regarding the task conditions and sensory modalities that determine the involvement of PPC. In Chapter 5, we examined the neural correlates of auditory and visual change detection in the PPC and tested the hypothesis that PPC is causally required to discriminate sensory modalities rather than to integrate them during an audiovisual change detection task.

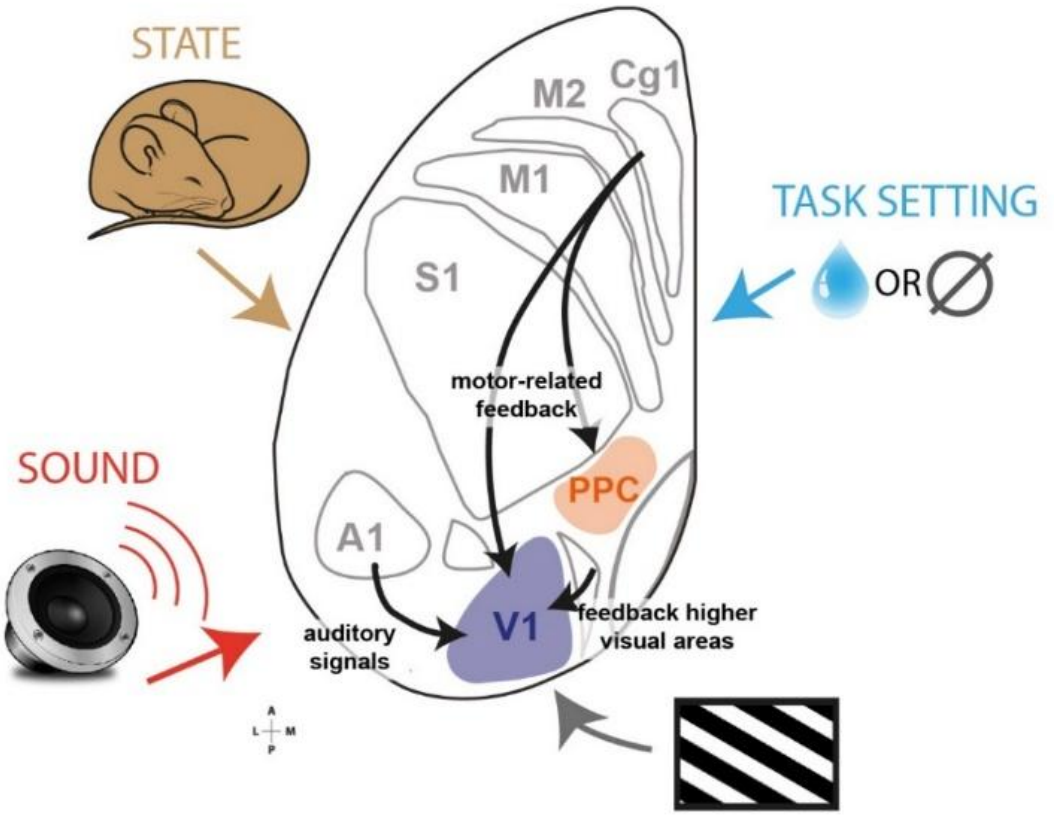


Figure 1.6: Contextual modulation of visual processing. *Processing of the same visual input (schematized as the oriented grating stimulus at the bottom right) was studied under conditions of variable states, sounds, and task settings. Black arrows denote relevant anatomical pathways known from the literature to modulate V1 and are hypothesized to play a role in the contextual feedback to V1. Note that context in this figure and the thesis is meant as sources of modulation external to V1.*

Chapter 2: Higher-order visual areas enhance stimulus responsiveness in mouse primary visual cortex

Matthijs N. oude Lohuis
Alexis Cerván Cantón
Cyriel M. A. Pennartz
Umberto Olcese

Published in:
Cerebral Cortex (2021)
10.1093/cercor/bhab414

Abstract

Over the past few years, the various areas that surround the primary visual cortex in the mouse have been associated with many functions, ranging from higher-order visual processing to decision making. Recently, some studies have shown that higher-order visual areas influence the activity of the primary visual cortex, refining its processing capabilities. Here we studied how *in vivo* optogenetic inactivation of two higher-order visual areas with different functional properties affects responses evoked by moving bars in the primary visual cortex. In contrast with the prevailing view, our results demonstrate that distinct higher-order visual areas similarly modulate early visual processing. In particular, these areas enhance stimulus responsiveness in the primary visual cortex, by more strongly amplifying weaker compared to stronger sensory-evoked responses (for instance specifically amplifying responses to stimuli not moving along the direction preferred by individual neurons) and by facilitating responses to stimuli entering the receptive field of single neurons. Such enhancement, however, comes at the expense of orientation and direction selectivity, which increased when the selected higher-order visual areas were inactivated. Thus, feedback from higher-order visual areas selectively amplifies weak sensory-evoked V1 responses, which may enable more robust processing of visual stimuli.

Introduction

The various areas which make up the mouse visual cortical system have, over the past decade, emerged as a prime model to study the functional architecture underlying vision in mammals (Andermann et al., 2011; Glickfeld and Olsen, 2017; Glickfeld et al., 2014; Marshel et al., 2011; Wang and Burkhalter, 2007). The anterior and lateral borders of primary visual cortex V1 are surrounded by an array of areas, collectively called higher-order visual areas (HVAs), each having distinct connectivity patterns and visual response properties, and providing a specific contribution to visual processing (Andermann et al., 2011; Marshel et al., 2011; Wang et al., 2012). Several studies have investigated what functions each of these areas might fulfill in visual processing. A wide range of functions has been found, complementing V1 in orientation discrimination and contrast detection (Jin and Glickfeld, 2020), spatial integration (Murgas et al., 2020), perception of higher-order visual features (Khastkhodaei et al., 2016) and illusory contours (Pak et al., 2020). Furthermore, HVAs partially overlap with the rodent posterior parietal cortex, and have been implicated in several functions beyond simple visual processing, for instance multisensory integration (Meijer et al., 2020; Olcese et al., 2013; Song et al., 2017), (multi)sensory evidence accumulation and decision making (Erlich et al., 2015; Hanks et al., 2015; Licata et al., 2017; Raposo et al., 2014) and navigation (Harvey et al., 2012; Krumin et al., 2018). Moreover, HVAs play a significant role in sensory processing by means of the input they provide not only to each other, but also to V1 (Wang et al., 2012).

Feedback projections from HVAs to V1 have been found to be functionally organized (Kim et al., 2018; Marques et al., 2018), similarly to local connections (Cossell et al., 2015; Ko et al., 2011) and feedforward projections from V1 to HVAs (Berezovskii et al., 2011; Glickfeld et al., 2013a). These feedback projections have been associated with a variety of essential forms of visual processing: response facilitation (Nurminen et al., 2018; Pafundo et al., 2016), surround suppression (Nassi et al., 2013; Nurminen et al., 2018; Vangeneugden et al., 2019) and predictive processing (Keller et al., 2020b). A recent study, in particular, showed that each HVA differently impacts the activity of V1 neurons based on their visual response properties (Huh et al., 2018). Inactivating either the anterolateral (AL) or posteromedial (PM) area primarily reduced responses of those V1 neurons showing functional properties similar to those of AL and PM, respectively. The Huh et al. (2018) study focused on tuning of V1 cells to spatial frequency and investigated how inactivating AL and PM modulates firing rate responses to drifting gratings moving along the preferred orientation of single neurons. Overall, previous studies thus indicate that feedback projections from HVAs to V1 may provide a mechanism to enhance processing of specific visual stimuli, based on the response properties of each HVA. We expanded the results of previous literature by combining optogenetics and ensemble recordings to investigate how HVAs contribute to a broad spectrum of V1 functions – such as orientation and direction selectivity, receptive field size and single-trial encoding of visual features – as a function of the speed of visual stimuli. We further compared anesthetized and awake conditions. Unconscious brain states have been associated with lacking or diminished recurrent processing (Keller et al., 2020b; Lamme et al., 1998; Makino and Komiyama, 2015) and we investigated whether functionally specific feedback is degraded under anesthetized conditions. Surprisingly, we found that, in addition to the previously reported, functionally specific feedback (in which modulation of V1 varies based on the functional tuning of each HVA), AL and PM similarly enhance V1 responsiveness to visual stimuli, during both wakefulness and anesthesia. Such enhancement is especially prominent for weak sensory-

evoked responses and for responses entering the receptive field of single neurons, but comes at the expense of orientation and direction selectivity. Thus, in addition to previously discovered functions, HVAs also contribute to amplifying V1 responses, especially to stimuli which would otherwise evoke small changes in spiking activity.

Materials and Methods

Subjects

All animal experiments were performed according to the national and institutional regulations. The experimental protocol was approved by the Dutch Commission for Animal Experiments and by the Animal Welfare Body of the University of Amsterdam. A total of 14 male mice from two transgenic mouse lines were used: PVcre (B6;129P2-Pvalbtm1(cre)Arbr/J, JAX mouse number 008069) and F1 offspring of this same PVcre line with Ai9-TdTomato cre reporter mice (B6.Cg-*Gt(ROSA)26Sor*$^{tm9(CAG\text{-}tdTomato)Hze}$/J, JAX mouse number 007909). Animals were at least 8 weeks of age at the start of experiments. Mice were group housed, with ad libitum access to water and food, under a reversed day-night schedule (lights were switched off at 8:00 and back on at 20:00). All experimental procedures were performed during the dark period.

Experimental Design

Headbar implantation

Mice were subcutaneously injected with the analgesic buprenorphine (0.025 mg/kg) and maintained under isoflurane anesthesia (induction at 3%, maintenance at 1.5–2%) during surgery. The skin above the skull was epilated, disinfected, and a circular area was removed with the edges glued to the outer parts of the skull using tissue adhesive (3M Vetbond, MN, United States) to prevent post-surgical infections. A custom-made titanium head-bar with a circular recording chamber (inner diameter: 5 mm) was positioned over the exposed skull of the left hemisphere to include visual, auditory and somatosensory cortices and attached using cyanoacrylate and C&B Super-Bond (Sun Medical, Japan).

Intrinsic optical imaging

To localize individual higher visual cortical areas we performed intrinsic optical imaging (IOI) under lightly anesthetized conditions (0.7-1.2% isoflurane). A vasculature image was acquired under 540 nm light before starting the imaging session. During IOI, the cortex was illuminated with monochromatic 630 nm light. Images were acquired at 1 Hz using an Adimec 1000m CCD camera (1004 x 1004 pixels) connected to a frame grabber (Imager 3001, Optical Imaging Inc, Germantown, NY, USA), defocused about 500-600 μm below the pial surface.

We presented visual, auditory and tactile stimuli. Visual stimuli consisted of full field drifting gratings (spatial frequency 0.05 cpd, temporal frequency 1.5 Hz) for 1 second in each of 8 directions. Auditory stimuli consisted of alternations between chirps sweeping up or down in frequency (1-40 kHz) and band-passed white-noise (1-40kHz) calibrated at 70 dB Sound Pressure Level (SPL). Tactile stimuli were full whisker-pad deflections driven by a piezo-actuator (18˚ angle). For each type of stimulation we acquired 8 seconds of baseline signal and 8 seconds of hemodynamic response during stimulation. The acquired frames

during the response were baseline-subtracted, averaged and thresholded to produce a map of localized individual primary and higher order areas. PM and AL were identified based on the IOI signal map in combination with previously published maps (Glickfeld and Olsen, 2017; Olcese et al., 2013; Wang and Burkhalter, 2007) and marked on the skull based on the vasculature image. After IOI, the recording chamber was covered with silicon elastomer (Picodent Twinsil) and mice were allowed to recover for 2-7 days.

Viral injections

Mice were subcutaneously injected with the analgesic buprenorphine (0.025 mg/kg) and maintained under isoflurane anesthesia (induction at 3%, maintenance at 1.5–2%) during surgery. We performed a small craniotomy over the area of interest (either PM or AL in distinct mice, identified using IOI) using a dental drill and inserted a glass pipette backfilled with AAV2.1-EF1a-double floxed-hChR2(H134R)-EYFP-WPRE-HGHpA (titer: 7×10^{12} vg/ml, 20298-AAV1, Addgene). Four injections of 13.8 nl were made at two depths (two at 700 µm and two at 400 µm below the dura) using a Nanoject pressure injection system (Drummond Scientific Company, USA). Each injection was spaced apart by at least 5 minutes from the next one to promote diffusion and prevent backflow. After viral injections, the recording chamber was covered with silicon elastomer (Picodent Twinsil) and mice were allowed to recover. In total, we performed successful injections in PM in 5 mice, and in AL in 4 mice.

Craniotomy

After at least 3 weeks to allow for robust viral expression, mice were subcutaneously injected with the analgesic buprenorphine (0.025 mg/kg) and maintained under isoflurane anesthesia (induction at 3%, maintenance at 1.5–2%) during surgery. We performed small (200 µm) craniotomies over the areas of interest (V1 and either PM or AL) using a dental drill. The dura was left intact if possible. The recording chamber was sealed off with silicon elastomer and the mice were allowed to recover for 24h.

In vivo electrophysiology

Mice were fixated in a custom-built holder in a dark and sound-attenuated cabinet. The body of the mouse was put in a tube (diameter: 4 cm) to limit body movements. The headbar was attached to a custom-made holder via two screws. Before recording sessions, mice were habituated to this type of head-fixation by daily progressive incremental time spent in head-fixation.

Recordings were performed either in an awake or anesthetized state and the order was counterbalanced across recording days. Under anesthesia, pure oxygen with isoflurane (at 0.6-1.2%) was delivered at 0.8 l/min. The level of anesthesia was monitored by observing breathing rate and neural activity. Isoflurane levels were slowly lowered over the course of a recording session to counteract tissue build-up and maintain a stable depth of anesthesia. Body temperature was monitored throughout and kept at 37.5 ˚C.

Extracellular recordings were performed with 32- or 64-channel microelectrode arrays (NeuroNexus, Ann Arbor, MI –A1x32-Poly2-10mm-50s-177, A4x8-5mm-100-200-177, or A1x64-Poly2-6mm-23s-160). Each recording session the electrode arrays were slowly inserted until the recording sites spanned the cortical layers. We verified visual responsiveness by displaying full-field gratings and reinserted the electrodes if there was

no robust visual responsiveness in neural activity. The number of recording sessions was limited to 3 to minimize recording from a damaged circuit. For some recording sessions electrodes were dipped in DiI (ThermoFisher Scientific) allowing better post hoc visualization of the electrode tract (Fig. 2.1C). After insertion, the exposed cortex and skull were covered with 1.3-1.5% agarose in artificial CSF to prevent drying and to help maintain mechanical stability. The ground was connected to the headbar and the reference electrode to the agarose solution. Recordings started at least 15 minutes after insertion to allow for tissue stabilization. Neurophysiological signals were amplified (x1000), bandpass filtered (0.1 Hz to 9 kHz) and acquired continuously at 32 kHz with a Digital Lynx 128 channel system (Neuralynx, Bozeman, MT).

Optogenetics

To locally photostimulate PM or AL, a 473 nm laser (Eksma Optics, Vilnius, Lithuania, DPSS 473nm H300) was connected with a fiber-optic patch cord to a fiber-optic cannula (ID 200 um, NA 0.48, DORIC lenses) that was positioned directly over the thinned skull at the area of interest. Photostimulation consisted of 10 ms pulses delivered at 20 Hz for the duration of visual stimulus presentation. Stimulus duration varied depending on the traversal time of the bar across the screen and, depending on travelled distance and speed of the bar, ranged from 0.45 s (vertical bar moving at 70˚/s) to 6.3 s (diagonal bar moving at 20˚/s). Light delivery was controlled by a shutter (Vincent Associates LS6 Uniblitz). During each session we simultaneously performed extracellular recordings in the areas of interest (V1 and either PM or AL) and adjusted laser power to the minimum that maximally inhibited neural activity (range: 2-15 mW total power).

Visual stimulation

Visual stimuli were gamma-corrected and presented with a 60 Hz refresh rate on an 18.5 inch monitor positioned at a 45˚ angle with the body axis from the mouse at 21 cm from the eyes, subtending 91 ˚ horizontally and 60˚ vertically. Three sets of visual stimuli were used.

Checkerboards – Before each session, we displayed full-field contrast-reversing checkerboards (full contrast, spatial frequency = 10 retinal degrees, temporal frequency of contrast reversal = 0.5 Hz, n=10 reversals) to estimate laminar electrode positioning (see below).

Bars – Each bar stimulus consisted of a single white bar (luminance = 133 cd/m^2) drifting across an isoluminant gray screen (luminance = 32 cd/m^2) in one of 8 directions at one of three speeds (20˚/s, 40˚/s or 70˚/s) either in absence or presence of photostimulation. Stimuli were separated by an inter-trial interval of 3 seconds and repeated 20 times. The total trial set therefore consisted of 8 (orientations) x 3 (speeds) x 2 (photostimulation conditions) x 20 (repetitions) = 960 trials.

Gratings – Grating stimuli consisted of full-field drifting square-wave gratings (70% contrast) for 2 seconds, separated by 2 seconds inter-trial interval. Similar to the bar stimuli, gratings drifted in one of 8 directions at one of three speeds (20˚/s, 40˚/s or 70˚/s) either in absence or presence of photostimulation for 20 repetitions. The three speeds were constructed based on combinations of spatial and temporal frequencies to optimize V1, PM and AL responsiveness (Andermann et al., 2011; Marshel et al., 2011): Slow 20˚/s: Spatial

frequency = 0.1 cpd, Temporal frequency = 2 Hz, Mid 40˚/s: Spatial frequency = 0.075 cpd, Temporal frequency = 3 Hz, Fast 70˚/s: Spatial frequency = 0.057 cpd, Temporal frequency = 4 Hz.

Histology

At the end of the experiment, mice were overdosed with pentobarbital and perfused with 4% paraformaldehyde in phosphate-buffered saline, and their brains were recovered for histology. We cut coronal 50 µm sections with a vibratome, stained them with DAPI (0.3 µM) and imaged mounted sections to verify the viral expression and recording sites. The borders of individual higher visual areas in individual animals are not definable based on an atlas. However, with this consideration in mind, data from five animals was excluded based on weak expression in putative PM or AL or strong off-target expression beyond PM or AL or into V1.

Data analysis

Spike sorting

Before spike detection the median of the raw trace of nearby channels (400<um) was subtracted to remove artefacts. Spike detection and sorting were done using Klusta and manual curation using the Phy GUI (Rossant et al., 2016). During manual curation each proposed single unit was inspected based on its waveform, autocorrelation function and firing pattern across channels and time. Only high-quality single units were included that (1) had an isolation distance higher than 10 (Schmitzer-Torbert et al., 2005), (2) had less than 0.1% of their spikes within the refractory period of 1.5 ms, (3) were present throughout the session.

Classification of neuron subtypes

Putative pyramidal and putative fast-spiking interneurons were separated based on the peak-to-trough delay of their average normalized action potential waveform (Niell and Stryker, 2008). The peak-to-trough delay was computed as the time between peak positive and peak negative voltage deflection (in ms) and single units with a delay lower than 0.45 ms were classified as narrow-spiking, while units with a delay higher than 0.55 ms were classified as broad-spiking. The rest remained unclassified. In total 76.9% were labeled as broad spiking, 20.6% as narrow-spiking and 2.5% as unclassified.

Laminar depth estimation

The laminar depth of each electrode was estimated based on current source density analysis (CSD) of the local field potential (LFP) in response to contrast-reversing checkerboard stimuli (see above). The CSD profile was computed by applying standard Nicholson-Freeman calculations on the low-pass filtered signal (<100 Hz, 4th order Butterworth filter) with Vaknin transform (Vaknin et al., 1988) with 0.4 Siemens per meter as conductivity. We calculated the CSD profile for each of the linear arrays of electrodes on our polytrode configuration separately and then merged the profiles. The electrode with the earliest visible sink was designated as the center of layer IV. Single units recorded from electrodes spanning 150 µm around this electrode were labeled as granular and units recorded from electrodes below and above this layer were labeled infra- and supragranular, respectively.

Firing rate response

To compute firing rates in response to visual stimuli, spikes times were aligned to stimulus onset, binned in 1 ms bins and convolved with a Gaussian window (50 ms standard deviation). To compute single trial responses for bar stimuli we first identified the time of peak response for each condition (orientation x speed) by averaging across trial repetitions without photostimulation. The response on each trial was obtained by averaging the single trial firing rate over 300 ms around this peak time (± 150 ms), after subtracting baseline response (firing rates computed, separately for each trial, in the [-2000, -200] ms window before stimulus onset). In a set of control analyses we averaged firing rates not over a fixed time window around peak time, but instead computed, as a function of bar speed, how much time a bar took to cover a certain portion of the visual field (5 or 10 deg, respectively). Time windows covering a receptive field of 5 deg corresponded to 500 ms at 20 deg/s, 250 ms at 40 deg/s and 142 ms at 70 deg/s (and twice as much for a 10 deg coverage). For grating stimuli the single-trial firing rate was averaged over 0-1000 ms after stimulus onset. Only neurons showing a significant sensory evoked response (defined as having an average z-scored response >1 for at least one bar direction) were retained for further analyses. Z-scoring was done by subtracting for each trial the mean firing rate of the baseline period (-1 to -0.2 seconds before stimulus) and dividing by the standard deviation of all baseline periods. The following table summarizes the number of V1 neurons that were retained for analysis:

Bar speed	20 deg/s	40 deg/s	70 deg/s
PM inactivation – awake recordings	207	206	185
AL inactivation – awake recordings	94	97	79
PM inactivation – anesthetized recordings	88	87	87
AL inactivation – anesthetized recordings	67	65	63

Table 2.1 : Number of recorded neurons.

Quantification of peak latency, tuning curves, and receptive field size

Peak response latency was defined as latency of the peak response (maximal z-scored firing rate) in the absence of optogenetic inactivation. This was determined for each direction and each speed separately and was determined only for those conditions in which a z-scored sensory-evoked response higher than 1 was found. Tuning curves for single neurons were quantified after computing, independently for each direction and separately for each speed, the firing rate response to a visual stimulus. The preferred orientation/direction was computed, separately for each speed, based on the bar direction eliciting the largest average firing rate response, in the absence of optogenetic inactivation. To align tuning curves, the preferred orientation of each neuron in the absence of optogenetic inactivation was set to 0 degrees and other orientations were displayed relative to this.

Receptive field size was computed separately for the average responses to each bar direction. We computed the response onset as the first time point after stimulus onset in which the z-scored firing rate response exceeded 1. The response offset was defined as the

first time point following response onset for which the z-scored firing rate response dropped below 1. The receptive field size for a given direction was computed as the duration of the response (time lag between response onset and offset) multiplied by the speed of the bar. Receptive field size was aligned to the preferred direction, as described for the tuning curves.

Orientation and direction selectivity

Orientation and direction selectivity were computed using a global orientation selectivity index (gOSI) and a global direction selectivity index (gDSI) (Ibrahim et al., 2016; Mariño et al., 2005; Ringach et al., 2002a). These two measures were computed as:

$$g \quad = \frac{\left\| \sum_{\theta} R(\theta) e^{2i\theta} \right\|}{\sum_{\theta} R(\theta)} \qquad \text{(Eq 2.1)}$$

and

$$g \quad = \frac{\left\| \sum_{\theta} R(\theta) e^{i\theta} \right\|}{\sum_{\theta} R(\theta)} \qquad \text{(Eq. 2.2)}$$

Here $R(\theta)$ is the baseline-corrected firing rate response of a neuron to a bar moving along direction θ and i is the imaginary unit. gOSI and gDSI vary between 0 and 1, with 0 indicating a neuron completely untuned for orientation/direction, and 1 a neuron only responding to a single orientation/direction, respectively. For the analysis of gOSI and gDSI, we only retained neurons that showed significant sensory-evoked responses to at least one direction in the non-opto condition. If a neuron did not respond (meaning that no action potential was fired) to any stimulus direction in the opto condition, its gOSI and gDSI values were undetermined. For this reason, such neurons were removed from further analysis about how optogenetic inactivation affects gOSI and gDSI values.

Decoding analysis

The population decoding analysis was done using a pseudo-population approach. Decoding was separately performed for awake and anesthetized recordings. All recorded neurons were pooled together (even if they were recorded in different sessions) and decoding was performed on a randomly selected number of neurons equal to the lowest available number of neurons per condition. For awake recordings, this amounted to 79 neurons; for anesthetized recordings, to 63 neurons (see the table above). In detail, we used the same number of neurons to decode the direction of a moving bar presented at every speed, without optogenetics, or with inactivation of either PM or AL; this procedure allowed us to fairly compare the different conditions (area being inactivated and bar speed). When pooling together data from different recording sessions, we only considered conditions (bars moving along a certain direction and speed) which had been repeated over at least 10 trials. For all conditions, 20 trials were sampled over recording sessions (with replacement, if fewer than 20 trials were present, and without replacement otherwise). This data was used to train a k-nearest neighbors classifier, which was trained to decode the direction of the bar being presented, based on the single-trial firing rate response (computed as described above as the average firing rate in a 300 ms window centered around the peak latency – here defined as latency of the peak response – of each neuron to a bar with a

certain speed and direction). The performance of the decoder was assessed with a leave-one-out cross-validation procedure. Training was repeated 100 times, each time with a different, random set of neurons and randomly sampled trials. For each training set, we computed the average decoding error (difference between the presented direction of a moving bar and the estimated direction). Different decoding approaches (random forest, support vector machine) did not yield significantly better results. In a set of control analyses, we also computed decoding accuracy as the proportion of trials in which the direction of movement of a bar was correctly estimated.

Statistical analysis

Statistical analyses were done using parametric methods (t-tests and ANOVAs) if the assumption of normality was not violated. This was verified via the use of a Kolmogorov-Smirnov test. Non-parametric tests were used otherwise. If applicable (i.e. when an Anova was performed), multiple comparisons were corrected using a Tukey post-hoc test. When multiple, independent comparisons were performed, p-values were corrected via the application of a Benjamini-Hockberg false discovery rate (FDR) procedure (Korthauer et al., 2019).

Data and software availability

Original data and the MATLAB, Python and R scripts used to perform the analyses presented in this manuscript are available by reasonable request to Umberto Olcese (u.olcese@uva.nl).

Results

To investigate how HVAs influence V1 responses, we focused on two areas with the largest known differences in tuning to spatial and temporal frequencies of visual stimuli: AL and PM. While AL neurons preferentially respond to visual stimuli with high temporal frequencies and low spatial frequencies, the opposite is true for area PM (Andermann et al., 2011; Marshel et al., 2011). We performed dual-area silicon probe recordings in head fixed mice from either V1 and AL or V1 and PM (Fig. 2.1A). Recordings were done in both the awake and anesthetized state. As recently reported (Keller et al., 2020b), feedback modulation from HVAs to V1 is reduced under anesthesia, but the effect of brain state on V1 response properties is poorly understood (Olcese et al., 2018), although previous studies reported a reduction in direction tuning in isoflurane anesthesia compared to wakefulness (Goltstein et al., 2015). Localized nano-injections of a viral vector mediating Cre-dependent expression of channelrhodopsin were performed in either AL or PM of PV-Cre mice (Madisen et al., 2012) – Fig. 1A,C. Areas V1, AL and PM were localized via intrinsic optical signal imaging (IOI, Fig. 2.1B). We verified that expression was confined to AL or PM and did not extend across the PM-V1 or AL-V1 borders (Fig. 2.1C, Supplementary Fig. 2.1A-G). Blue-light illumination was used to inactivate either area AL or PM, via over-activation of parvalbumin-positive (PV+) interneurons (Olcese et al., 2013) – Fig. 2.1D-G. We experimentally verified that optogenetic inactivation was confined to areas AL and PM, and did not affect V1 directly (Fig. 2.1C,H). Specifically, inactivation of either AL or PM greatly reduced the activity of putative excitatory neurons in the illuminated area (Fig. 2.1G), but only had a minor effect of spontaneous firing activity in V1 (Fig. 2.1H) or non-

photostimulated PM or AL (Supplementary Fig. 2.1H,I) – see also Supplementary Fig. 2.1K. Furthermore, positioning the fiber tip over an uninfected control area (primary somatosensory area) did not affect firing rates in V1, excluding the possibility that scattered light reached the retina and affected visual responses (Supplementary Fig. 2.1J). The effectiveness of photostimulation in the target area increased as a function of laser power (Supplementary Fig. 2.1L). Moving bars were used as visual stimuli to evoke activity in V1 and HVAs (Fig. 2.1I). Compared to drifting gratings, moving bars enable to assess receptive field sizes of the recorded neurons (Niell and Stryker, 2008). Bars were moving over 8 different orientations at three different speeds, namely those that previous studies (70 deg/s) (Andermann et al., 2011; Marshel et al., 2011). To verify this speed preference, we computed, separately for each neuron and area, the response to a bar moving along the preferred orientation, independently for each speed. For each neuron, responses were normalized to the speed evoking the strongest response (Fig. 2.1J). For recordings performed under anesthesia, the speed preference of neurons in each area was in line with the literature (Fig. 2.1J-right). During wakefulness we found a shift for all areas to lower preferred speeds compared to anesthesia (Fig. 2.1J-left), although the differential speed preference of neurons in area PM and AL – with area PM selective for low speeds and AL for faster speeds – was preserved. Optogenetic inactivation of either area PM or AL strongly reduced sensory-evoked responses in putative excitatory neurons in the illuminated area – PM or AL, respectively (Supplementary Fig. 2.2). indicated as being preferred by PM (20 deg/s), V1 (40 deg/s) and AL. Inactivation of single HVAs also reduced – to a lesser extent – sensory-evoked responses in V1 (Fig. 2.1K and Supplementary Fig. 2.3A-B, see also later sections). The reduction in V1 activity was, in contrast to the effect of the manipulation on PM and AL, limited to sensory-evoked responses, with only minor effects on spontaneous activity (Fig. 2.1G, Supplementary Fig. 2.1K), and can be interpreted in the first instance to be the consequence of impaired recurrent connectivity from AL or PM (see also Materials and Methods). We also tested whether temporally extended photostimulation during trials with low stimulus speeds might have an effect on V1 independent from optogenetics, but rather due to light-induced heating of the cortical tissue. Nevertheless, we found the effect of PM or AL inactivation on V1 to be independent of the duration of photostimulation (Supplementary Fig. 2.1M).

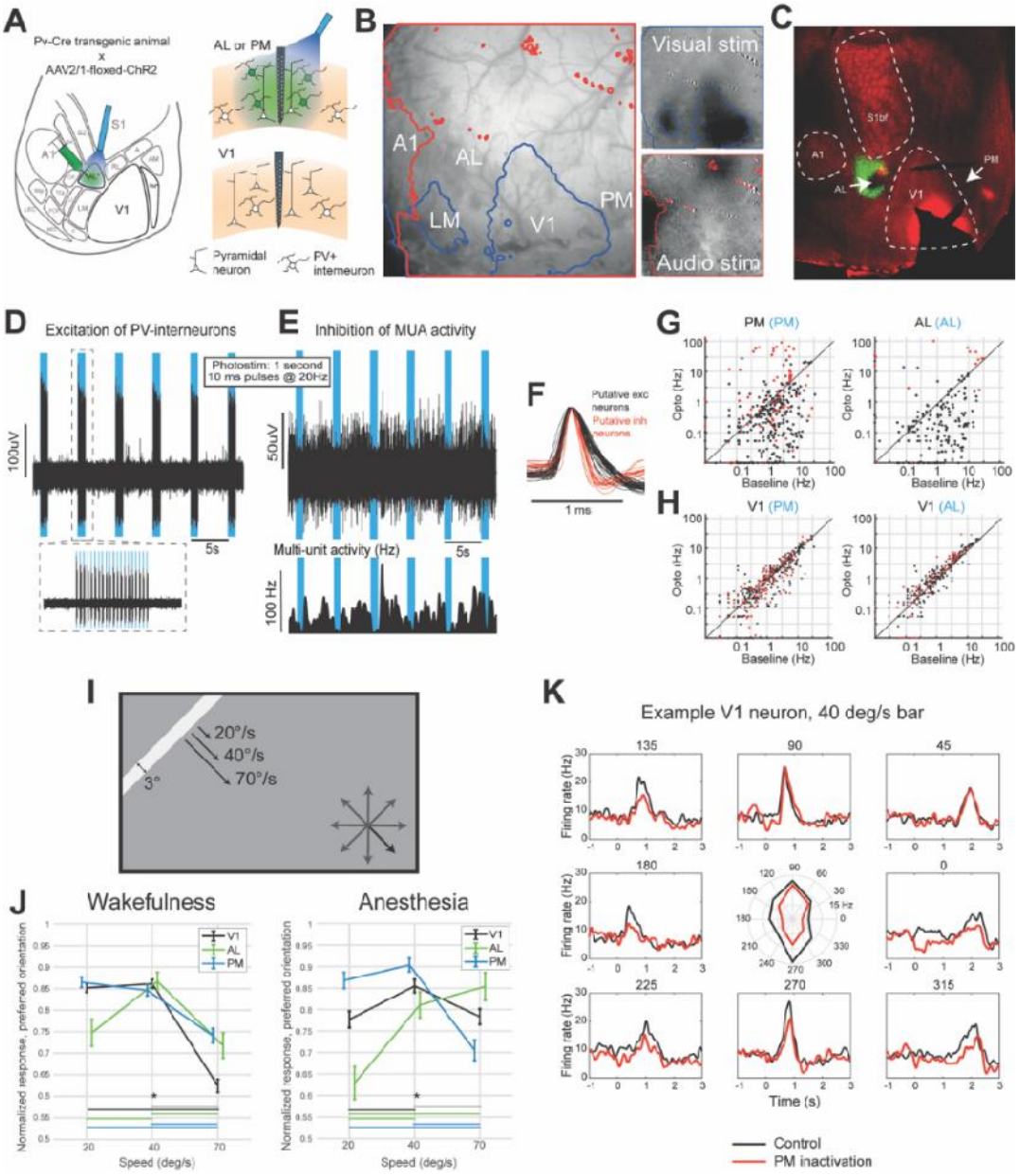


Figure 2.1 – Experimental setup. A. *Schematic of the experimental design. Left: top view of the left cortical hemisphere of a mouse, with subdivision in cortical areas – based on (Wang and Burkhalter, 2007). Adeno-associated viral vector mediating the Cre-dependent expression of ChR2 was injected in area AL (or PM, not shown). During experiments an optic fiber (blue) was placed on top of AL (or PM) to over-activate Cre-expressing PV+ interneurons and inactivate area AL (or PM). Right: scheme of coronal sections of either AL/PM (top) or V1 (bottom) showing laminar probe recordings in both areas. Expression of ChR2 and fiber-optic-mediated illumination were confined to area AL or PM (top).* ***B.*** *Intrinsic signal imaging was used to localize cortical areas. Visual stimuli (top right) and auditory stimuli*

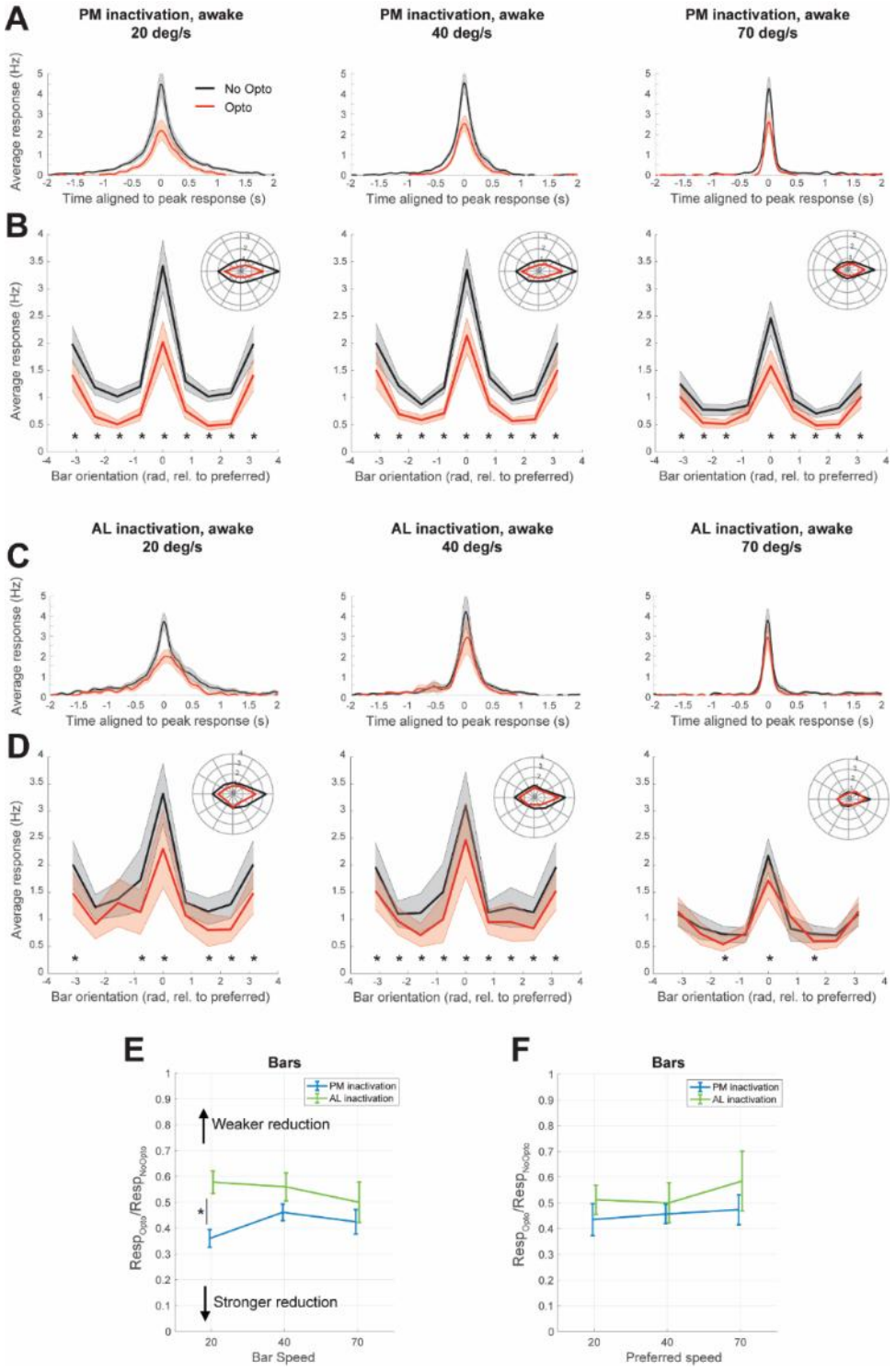
(bottom right) were used to activate and thus identify the location of visual and auditory cortices. The borders of visually- and auditory-evoked signals (blue and red curves, respectively) were overlaid on the vessel map (left) to identify the location of V1, AL and PM. In this example, V1, LM and A1 were directly activated by visual or auditory stimuli. The location of AL and PM was determined based on published maps of the mouse visual system (see A). ***C.*** *Coronal section showing ChR2-conjugated GFP expression in area AL (green). Red reflects both the tdTomato fluorescent protein expressed in PV+ interneurons as well as the location of laminar probes stained with DiI.* ***D.*** *Example neuronal trace from a PV+ interneuron recorded in area PM during optogenetic illumination in control trials (during awake recordings). Raw trace from a channel showing spiking activity evoked in a PV+ interneuron by optogenetic illumination. Blue areas indicate 1 s periods in which the blue laser was on. The inset shows the pattern of optogenetic illumination (10 ms ON – 40 ms OFF) during each illumination period.* ***E.*** *Top: Example multiunit activity (MUA) trace recorded in area PM during optogenetic illumination in control trials. In contrast with panel D, spiking activity decreased during illumination periods. Bottom: Firing rate traces as extracted from MUA activity shown above. Notice the decrease in firing rates during illumination.* ***F.*** *Average action potential waveforms from a selection of putative excitatory neurons (black, characterized by broad spikes) and putative inhibitory neurons (red, characterized by narrow spikes).* ***G.*** *Scatter plots of firing rates of individual neurons during spontaneous baseline activity and during optogenetic stimulation in areas PM (left) or AL (right, recording area in black font, photostimulated area in blue font between brackets), for both putative excitatory and inhibitory neurons (black and red points, respectively). Average spontaneous firing rates for putative excitatory neurons significantly decreased upon optogenetic stimulation of PM and AL in both areas (PM, n=338 neurons, mean values: 2.3 Hz and 1.0 Hz, $p=1.31*10^{-26}$; AL, n=199 neurons, mean values 2.5 Hz and 0.9 Hz, $p=2.06*10^{-21}$; Wilcoxon signed rank test), while those for putative inhibitory neurons increased significantly in AL (n=18 neurons, mean values 8.6 to 24.3 Hz, p=0.0256, Wilcoxon signed rank test), but not PM (n=72 neurons, mean values: 3.4 Hz and 13.1 Hz, p=0.08, Wilcoxon signed-rank test).* ***H.*** *Same as G, but for neurons recorded in V1 during optogenetic stimulation of either PM (left) or AL (right). Optogenetic stimulation of either PM or AL had minor but still significant effects on spontaneous activity in V1. Left: average spontaneous firing rates for V1 during optogenetic stimulation of PM significantly decreased for putative excitatory neurons: n=351 neurons, 3.2 Hz and 2.4 Hz, $p=3.9*10^{-11}$, and putative inhibitory neurons: n=121 neurons, 2.5 Hz to 2.1 Hz, p=0.019; Wilcoxon signed-rank test). Right: average spontaneous firing rates for V1 during optogenetic stimulation of AL significantly increased for putative excitatory neurons: n=283 neurons, 2.8 Hz and 2.9 Hz, p=0.041, and putative inhibitory neurons: n=76 neurons, 2.6 Hz to 3.4 Hz, p=0.006; Wilcoxon signed-rank test). Although spontaneous firing rates in V1 were affected by optogenetic stimulation in PM or AL, such increments or decrements were much weaker than those reported in PM and AL (see also Supplementary Fig. 2.1E).* ***I.*** *Outline of the visual stimuli (moving bars moving at different speed along 8 possible directions).* ***J.*** *Speed preference for neurons located in V1 (black), PM (blue) and AL (green) as a function of brain state (left: wakefulness; right: anesthesia). For each neuron, responses to the preferred orientation were computed across the three bar speeds, and normalized to the highest response (corresponding to the preferred bar speed). Asterisks indicate significant differences between speeds, for neurons located in the same area ($p<0.05$, one-way Anova with post-hoc Tukey test; V1-awake: $F=109.41$, $p=3.13x10^{-43}$; PM-awake: $F=18.5$, $p=1.8x10^{-8}$; AL-awake: $F=8.19$, $p=0.0004$; V1-anesthetized: $F=5.79$, $p=0.0033$; PM-anesthetized: $F=27.47$, $p=1.8x10^{-11}$; AL-anesthetized: $F=9.47$, $p=0.0001$). The number of neurons used for this analysis was: V1-awake: 339; V1-anesthetized: 158; PM-awake: 222; PM-anesthetized: 89; AL-awake: 70; AL-anesthetized: 48.* ***K.*** *PSTHs computed for an example neuron in V1, for bars moving at 40 deg/s across the 8 different orientations, with and without optogenetic stimulation of area PM (black and red traces, respectively). The polar plot at the center of the panel shows the tuning curve of the example neuron.*

Inactivation of areas AL and PM globally decreases V1 responses to moving bars

Having established that areas AL and PM show sensory-evoked responses which differ based on the specific speed at which presented bars move (Fig. 2.1J), we wondered whether inactivation of AL and PM would differentially modulate V1 responses to bars moving at different speeds. Surprisingly, we found that inactivating either AL or PM consistently reduced V1 responses to bars moving at all the speeds we tested, and for both preferred and non-preferred orientations (Fig. 2.2A-D, see also example traces of single neurons in

Fig. 2.1K and Supplementary Fig. 2.3A-B). Importantly, direction preference was not modified by inactivation of PM and AL (Supplementary Fig. 2.4). This was the case during both awake (Fig. 2.2A-D) and anesthetized (Supplementary Fig. 2.5) recordings. We next asked if the extent to which V1 responses were reduced varied as a function of bar speed. For each neuron and bar speed, we computed the relative change in the response to a bar moving along the neuron's preferred direction following optogenetic inactivation of either area AL or PM. No significant difference was found between speeds when inactivating either AL or PM (Fig. 2.2E). Only for bars moving at 20 deg/s, we found than PM inactivation reduced V1 responses more strongly than AL inactivation. To further explore the possible occurrence of a functionally specific effect, we subdivided V1 neurons in three groups based on the bar speed for which they showed the highest response – cf. (Huh et al., 2018) – and assessed whether AL and PM inactivation had different effects for the three groups of neurons. While we found no significant differences for bars (Fig. 2.2F), we did find some differences when presenting drifting gratings, in line with Huh et al. (2018) – see Supplementary Fig. 2.6.

We also wondered if the reported effects of optogenetic inactivation of PM and AL on V1 responses could be at least partially ascribed to characteristics of the measurements, or to the methods we use to quantify sensory-evoked responses. To address this, we first tested if the higher number of V1 neurons recorded during PM compared to AL inactivation might explain the different modulation of V1 responses that we observed for low speeds (Fig. 2.2E). We repeated the analysis of Fig. 2.2E by resampling V1 neurons recorded during PM inactivation such that the same number of neurons was analyzed as was recorded during AL inactivation (see the legend of Supplementary Fig. 2.7A for details). While, as expected, the variability of response modulation following PM inactivation increased when the number of neurons used in the analysis decreased (Supplementary Fig. 2.7A), the main result reported in Fig. 2.2E were preserved (Supplementary Fig. 2.7B, note the stronger reduction in responses following PM compared to AL inactivation only for bars moving at 20 deg/s). Furthermore, we tested if the results we obtained might depend on the temporal window used for quantifying sensory-evoked responses. Instead of a window of a fixed duration across speeds, we used a window of variable duration (still centered around peak responses), corresponding to the time a moving bar took to cover a visual angle of either 5 or 10 deg, irrespective of bar speed. In both cases, we confirm the results reported in Fig.


A
PM inactivation, awake
20 deg/s
PM inactivation, awake
40 deg/s
PM inactivation, awake
70 deg/s
No Opto
Opto
Average response (Hz)
Time aligned to peak response (s)
B
Average response (Hz)
Bar orientation (rad, rel. to preferred)
C
AL inactivation, awake
20 deg/s
AL inactivation, awake
40 deg/s
AL inactivation, awake
70 deg/s
Average response (Hz)
Time aligned to peak response (s)
D
Average response (Hz)
Bar orientation (rad, rel. to preferred)
E
Bars
PM inactivation
AL inactivation
Weaker reduction
Stronger reduction
Resp_Opto/Resp_NoOpto
Bar Speed
F
Bars
PM inactivation
AL inactivation
Resp_Opto/Resp_NoOpto
Preferred speed

(figure on the previous page) Figure 2.2 – Optogenetic inactivation of area AL and PM during wakefulness depresses sensory-evoked responses in V1. A. *PSTHs (aligned to peak latency in the absence of optogenetic inactivation) averaged over all V1 neurons responding to visual stimuli during awake recordings in the absence or presence of optogenetic inactivation of area PM (black and red curves, respectively). Curves with shading indicate mean ± SEM. From left to right: responses to bars moving at 20, 40 and 70 deg/s.* ***B.*** *Tuning curves of average responses of V1 neurons during awake recordings to bars moving at different orientation (aligned by the preferred orientation in the absence of optogenetic inactivation, which is set at 0 rad) in the absence or presence of optogenetic inactivation of area PM (black and red curves, respectively). Curves with shading indicate mean ± SEM. Inset: Average response displayed in polar coordinates. Asterisks indicate significant differences between responses to bars moving at a given orientation in the absence or presence of PM inactivation ($p<0.05$, paired t-test, FDR-corrected). From left to right: responses to bars moving at 20, 40 and 70 deg/s.* ***C.*** *Same as A, but now for the inactivation of area AL.* ***D.*** *Same as B, but now for the inactivation of area AL.* ***E.*** *Relative effect of optogenetic inactivation of either PM (blue) or AL (green) on V1 responses evoked by moving bars during awake recordings ($Resp_{Opto}/Resp_{NoOpto}$), as a function of bar speed. Asterisks indicate significant differences either between speeds (for a given area) or between inactivation of distinct areas, given the same speed ($p=0.0009$, 2-way Anova with post-hoc Tukey test: significant main effect for inactivated area ($F=4.83$, $p=0.028$)). Error bars indicate mean ± SEM.* ***F.*** *Same as E, but subdividing neurons based on their preferred speed, and only considering the effect of optogenetic inactivation on the preferred speed. The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*

2.2E (Supplementary Fig. 2.7C-D). Finally, we tested the use of geometric mean to compute across-neuron average modulation values. We observed similar results with respect to the use of arithmetic mean, but wider confidence intervals (cf. Fig 2.2E, Supplementary Fig. 2.7E).

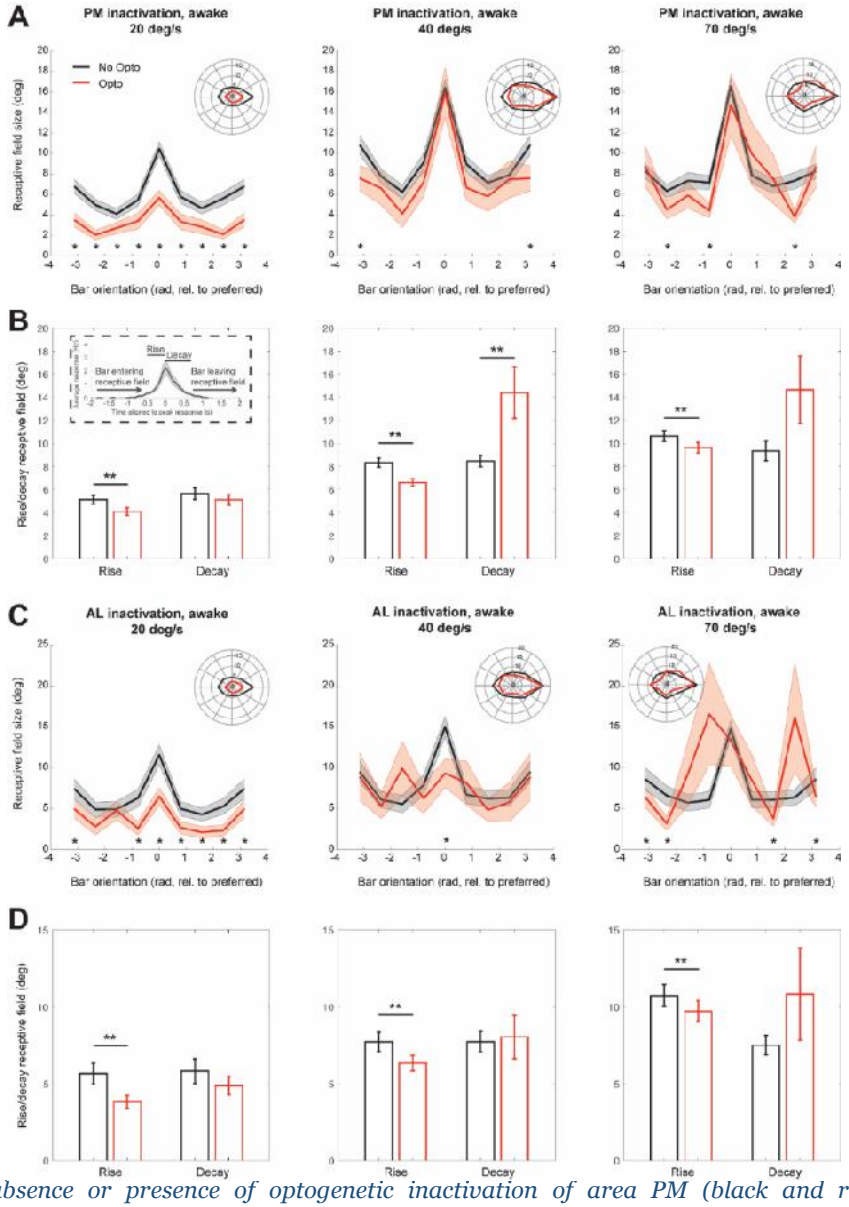
In conclusion, inactivating both AL and PM generally decreased V1 responses to moving bars. While we were able to confirm the previously reported presence of a functionally specific effect of AL and PM inactivation on V1 activity (i.e., being dependent on the speed preference of each HVA), this effect was weaker than the generalized decrease in responses observed across speeds (Fig. 2.2E-F, Supplementary Fig. 2.6).

Inactivation of AL and PM decreases responses to stimuli entering the receptive field of V1 neurons

The general reduction in visually evoked responses (Fig. 2.1K, 2.2A,C) made us wonder whether inactivating areas PM and AL would also reduce the receptive field size of V1 neurons. Since we used moving bars, any estimate of receptive field size computed through responses to such stimuli is conflated with response amplitude (unlike estimates made via the use of non-moving stimuli). Therefore, in line with the reduction in amplitude and duration of responses to moving bars (Fig. 2.2), in awake recordings we also observed a reduction of receptive field size, which was especially pronounced for bars moving at low speed (20 deg/s; Fig. 2.3A,C), but still present for bars moving at higher speeds, albeit only at some orientations (Fig. 2.3A,C). Nevertheless, the use of moving bars also allowed us to assess whether receptive field size (which, as we mentioned, goes in parallel with the size of sensory-evoked responses), is differentially affected by inactivation of PM and AL based on whether a stimulus is entering or leaving the receptive field of a neuron. This is an

important question in view of the possible role of HVAs in predictive processing (Keller et al., 2020b).

Figure 2.3 – Optogenetic inactivation of area AL and PM during wakefulness asymmetrically reduces receptive field size of V1 neurons in the rising but not in the decaying phase of sensory evoked responses. A. *Tuning curves of receptive field size of V1 neurons during awake recordings to bars moving at variable orientation (aligned by the preferred orientation, which is set at 0 rad)*



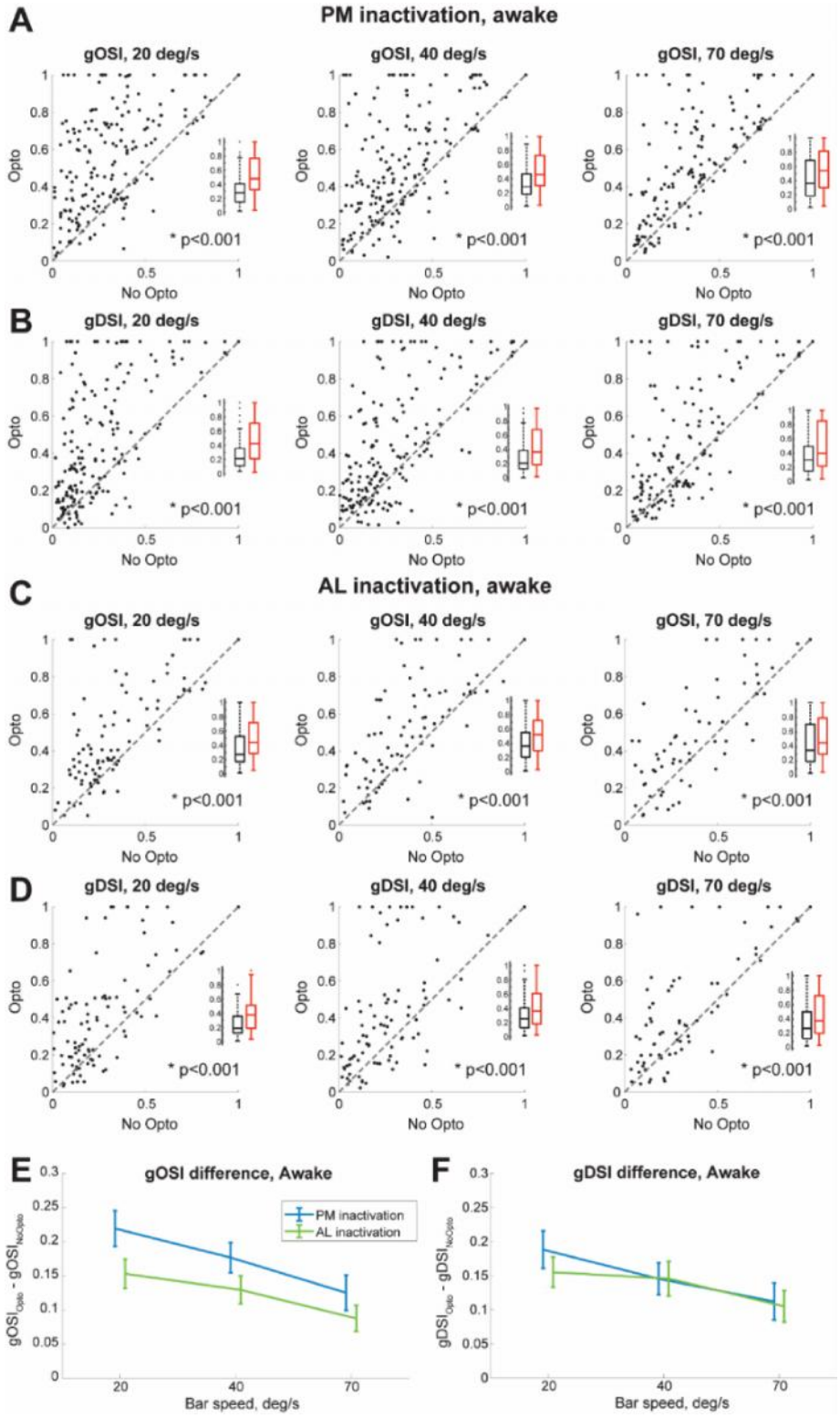
in the absence or presence of optogenetic inactivation of area PM (black and red curves, respectively). Curves with shading indicate mean ± SEM. Inset: average receptive field size displayed in polar coordinates; preferred orientation is aligned to the 0 deg (rightward) direction. Asterisks indicate significant differences between receptive field size to bars moving at a given

orientation in the absence or presence of optogenetic inactivation of area PM (p<0.05, paired t-test, FDR-corrected). From left to right: responses to bars moving at 20, 40 and 70 deg/s. ***B.*** *Receptive field size computed in the rising or decaying phase of sensory-evoked responses to moving bars moving along the direction preferred by individual neurons (from stimulus onset to peak response and after peak response, respectively) in the absence or presence of optogenetic inactivation of area PM (black and red bars, respectively). Bars indicate mean ± SEM. Asterisks indicate significant differences between receptive field sizes measured in the absence or presence of optogenetic inactivation of area PM (paired t-test; *: p<0.05, **: p<0.01). From left to right: responses to bars moving at 20, 40 and 70 deg/s. (P-values for significant differences: 20 deg/s rising phase p=1.2x10^{-6}; 40 deg/s rising phase p=6.8x10^{-10}; 70 deg/s rising phase p=3.1x10^{-6}; 40 deg/s decay phase p=0.008). The inset shows an example PSTH (same as Fig. 2.2A) to graphically explain what the rising and decaying phases correspond to.* ***C.*** *Same as A, but now for optogenetic inactivation of area AL.* ***D.*** *Same as A, but now for optogenetic inactivation of area AL. (P-values for significant differences: 20 deg/s rising phase p=0.0003; 40 deg/s rising phase p=0.0007; 70 deg/s rising phase p=0.0037). The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*

To address this, we separately computed the receptive field size in the rising phase of the evoked response (from response onset to peak response) compared to the decaying phase (after peak response) – see the inset in Fig. 2.3B. We performed this analysis for responses to stimuli moving along each neuron's preferred direction. After inactivation of either area PM or AL, responses during the rising phase became more spatially localized, i.e. neurons start responding later to moving bars entering the receptive field compared to control conditions (Fig. 2.3B,D). Receptive field sizes during the decay phase (i.e. for bars leaving the receptive field) were generally unchanged, except for an increase for bars moving at 40 deg/s upon PM inactivation (Fig. 2.3B,D). The preferential effect of HVA inactivation on the rising phase of the response to the preferred orientation can also be observed in the single neuron examples in Fig. 2.1K and Supplementary Fig. 2.3A. Results were very similar for recordings performed under anesthesia, with the main difference being larger receptive fields for V1 neurons in anesthetized than in awake recordings (Supplementary Fig. 2.8). Thus, inactivating AL and PM specifically reduces and delays V1 responses to moving bars entering the receptive field of V1 neurons, in line with a role of HVAs in generating predictions about upcoming visual stimuli.

Orientation and direction selectivity of V1 neurons are enhanced for moving bars when AL and PM are inactivated

To better explore the functional significance of the reduction in V1 responses and receptive field size, we wondered how inactivating AL and PM might affect orientation and direction selectivity of V1 neurons. These were quantified by computing, respectively, a global


A
PM inactivation, awake
gOSI, 20 deg/s
gOSI, 40 deg/s
gOSI, 70 deg/s
B
gDSI, 20 deg/s
gDSI, 40 deg/s
gDSI, 70 deg/s
C
AL inactivation, awake
gOSI, 20 deg/s
gOSI, 40 deg/s
gOSI, 70 deg/s
D
gDSI, 20 deg/s
gDSI, 40 deg/s
gDSI, 70 deg/s
Opto
No Opto
* p<0.001
E
gOSI difference, Awake
PM inactivation
AL inactivation
gOSI Opto - gOSI NoOpto
Bar speed, deg/s
F
gDSI difference, Awake
gDSI Opto - gDSI NoOpto
Bar speed, deg/s

(figure on the previous page) Figure 2.4 – Optogenetic inactivation of area AL and PM during wakefulness enhances orientation and direction selectivity in V1. A. *Scatter plots showing the orientation selectivity of V1 neurons during awake recordings in the absence or presence of optogenetic inactivation of area PM (x and y axis, respectively). Each point corresponds to a single neuron. Asterisks indicate significant differences between gOSI for bars moving at a given orientation in the absence or presence of optogenetic inactivation of area PM (see exact p value in each panel, Wilcoxon signed rank test). From left to right: gOSI for bars moving at 20, 40 and 70 deg/s. In each panel, boxplots depict descriptive statistics for gOSI values in No-Opto and Opto conditions (black and red, respectively).* ***B.*** *Same as A, but now for direction selectivity (gDSI).* ***C.*** *Same as A, but now for optogenetic inactivation of area AL.* ***D.*** *Same as B, but now for optogenetic inactivation of area AL.* ***E.*** *Average change in the orientation selectivity of V1 neurons during awake recordings as a function of bar speed and area being inactivated (PM: blue; AL: green). Error bars indicate mean ± SEM. No significant difference was found, neither between bar speeds, nor between areas.* ***F.*** *Same as E, but now for direction selectivity. No significant difference was found, neither between bar speeds, nor between areas. The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*

orientation selectivity index (gOSI) and a global direction selectivity index (gDSI) (Ibrahim et al., 2016; Mariño et al., 2005; Ringach et al., 2002a) – see Materials and Methods. Values of gOSI and gDSI were in line with those previously reported for mouse V1 (Ibrahim et al., 2016). Based on previous literature (Pafundo et al., 2016), we hypothesized that inactivation of HVAs would reduce both gOSI and gDSI. To our surprise, both gOSI and gDSI instead increased, across all bar speeds and in both wakefulness and anesthesia, regardless of whether AL or PM was inactivated (Fig. 2.4A-D, Supplementary Figure 2.9). No significant difference was found between inactivation of AL or PM, or between bar speeds (Fig. 2.4E-F) – although a more prominent enhancement for both gOSI and gDSI was observed at low speeds compared to high speeds when PM was inactivated during anesthesia (Supplementary Fig. 2.9). The increase in orientation and direction selectivity was observable in single neurons (Fig. 2.1K, Supplementary Fig. 2.3A-B) and in peak-normalized tuning curves (Supplementary Fig. 2.3C-D). Overall, these results suggest that, in contrast with previous literature (Pafundo et al., 2016), inactivation of AL and PM differentially reduces responses to bars moving along preferred and non-preferred orientations, in a way that enhances orientation and direction selectivity.

Single-trial decoding of the orientation of moving bars improves in V1 during wakefulness following inactivation of AL or PM

Orientation and direction selectivity indices are computed over average responses to visual stimuli. Therefore, we wondered if, in spite of enhancing gOSI and gDSI, inactivating AL and PM might have a different effect at the single-trial level. We reasoned that a reduction in the amplitude of sensory evoked responses might also reduce response variability at the average level. Thus, the improved gOSI and gDSI that we reported could be the consequence of both a reduction in response variability as well as of differential changes in sensory evoked responses to stimuli moving along preferred vs. non-preferred orientations. On the other hand, single-trial response selectivity would directly reflect changes in sensory evoked responses and not a reduction in across-trial variability. Thus, we implemented a pseudopopulation-based decoding approach (i.e. performed by pooling together neurons recorded in different sessions and animals) to measure how well the direction of moving bars could be decoded from single-trial V1 responses (see Materials and Methods). In line with the increase in gOSI and gDSI, we found that, during awake recordings, inactivating AL or PM significantly enhanced single-trial decoding of bar orientation, irrespective of bar

speed, and with a stronger improvement following AL than PM inactivation (Fig. 2.5). Notably, decoding error in control conditions was relatively high (about 70 deg), but decoding performance was comparable to recent literature (Cai et al., 2018) when quantified as the percentage of correctly classified directions (Supplementary Fig. 2.10A). Importantly, decoding performance vastly improved upon inactivation of PM or AL (40 deg). Similar results were observed for recordings under anesthesia, although the improvement in decoding following inactivation of PM or AL was smaller than for awake recordings (Supplementary Fig. 2.10B). Thus, inactivation of AL and PM not only reduces visually evoked responses and makes receptive fields smaller (i.e., spatially more precise), but also – possibly by differentially affecting responses to bars moving along preferred vs. non-preferred orientations and directions – enhances the selectivity of V1 neurons to the orientation and direction of moving stimuli, both at the average and single-trial level.

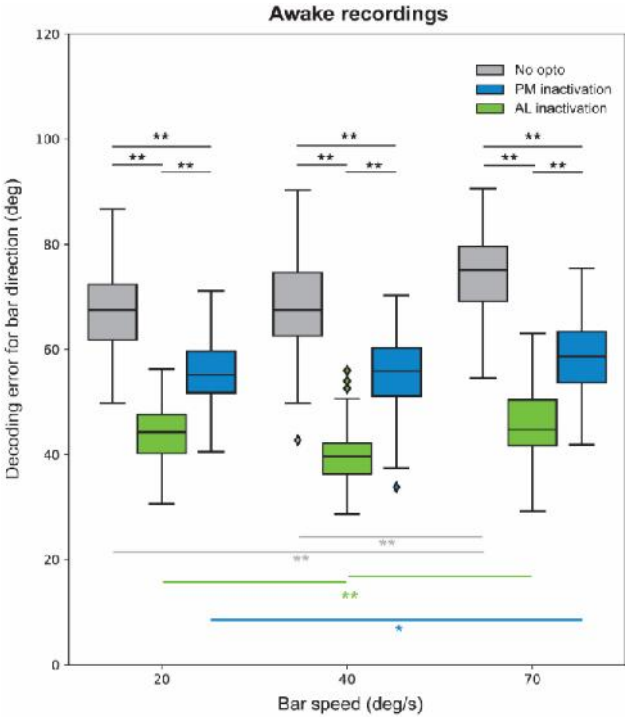


Figure 2.5 – Optogenetic inactivation of area AL and PM during wakefulness enhances single-trial decoding of bar orientation in V1. *Boxplots show the error (in deg) made by a decoder trained to estimate the orientation of a moving bar presented during individual trials, from the activity of a pseudo-population of V1 neurons (see Materials and Methods for decoding analysis). A lower decoding error indicates better ability to decode the orientation of a moving bar. Boxplots were separately computed for trials without optogenetic stimulation (grey) and for trials in which either AL (green) or PM (blue) was inactivated. Asterisks indicate significant differences (*: p<0.05, **: p<0.001; 2-way Anova with post-hoc Tukey test (main effect for inactivacted area: F=1056.6, p=0.0000; main effect for bar speed: F=56.6, p=0.0000; interaction effect: F=3.7, p=0.0052). The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*

AL and PM provide a modulatory gain to V1, which enhances weak visual responses during wakefulness

The above results paint a counter-intuitive role of AL and PM, which would have a primarily detrimental role on the response selectivity of V1 neurons if their sole purpose was to signal the orientation and direction of visual stimuli. We thus investigated in more depth whether inactivating AL and PM has a differential effect based on the size of a neuron's response to a given orientation. We reasoned that the modulation provided by AL and PM onto V1 might differentially enhance responses of different sizes (such as those elicited by bars moving along preferred vs. non-preferred orientations and directions), as previously shown for other forms of cortical gain modulation (Ferguson and Cardin, 2020). If the role of AL and PM is to selectively enhance small V1 responses, to enhance the reliability of responses to – for instance – small sensory

stimuli just entering the receptive field of a neuron, this might also explain the depressed orientation and direction selectivity which PM and AL induce. To test this, we first assessed whether AL and PM might provide a form of additive or multiplicative gain modulation of V1 responses, as these are often found in visual cortex (Wilson et al., 2012). To this aim, we subtracted or divided V1 responses following optogenetic inactivation of AL and PM by V1 responses in control conditions (Fig. 2.6A-B). During awake recordings, AL and PM implemented neither an additive nor a multiplicative form of gain modulation. In case of additive modulation, the curves in Fig. 2.6A would have been flat, reflecting a similar reduction in evoked responses following inactivation of AL or PM; similarly, in case of multiplicative modulation, the curves in Fig. 2.6B would not have shown differences between preferred and non-preferred bar directions. Conversely, inactivating AL and PM reduced the responses of V1 neurons to their preferred orientation significantly more than (weaker) responses to non-preferred orientations in terms of absolute difference (Fig. 2.6A), thus indicating a non-additive form of modulation. Similarly, in terms of relative amplification (division between responses with or without optogenetic inactivation of AL or PM), responses to preferred orientations were reduced less than smaller responses to non-preferred orientations (Fig. 2.6B). Next, we computed how optogenetically-induced response difference ($Resp_{NoOpto}$-$Resp_{Opto}$) and response amplification ($Resp_{Opto}/Resp_{NoOpto}$) of V1 sensory-evoked responses vary as a function of the amplitude of sensory-evoked responses (irrespective of the orientation/direction of the moving bar to which a response is made). For the inactivation of both PM and AL, V1 responses were modulated in a way that does not conform to either an additive or multiplicative model (cf. the colored and grey lines in Fig. 2.6C,D). Low-amplitude responses were more strongly suppressed than high-amplitude one, with a cutoff present at around 3 Hz. (Fig. 2.6C-D), which is in line with our earlier analyses about the modulation of responses to preferred vs. non-preferred directions. The non-uniformity of response modulation was significant in all instances ($p<2x10^{-16}$, Chi-square goodness of fit test against a uniform distribution) and similar results were found under anesthesia (Supplementary Fig. 2.11). Altogether, these results show that AL and PM provide a form of gain modulation that selectively enhances weak responses of V1 neurons, such as those

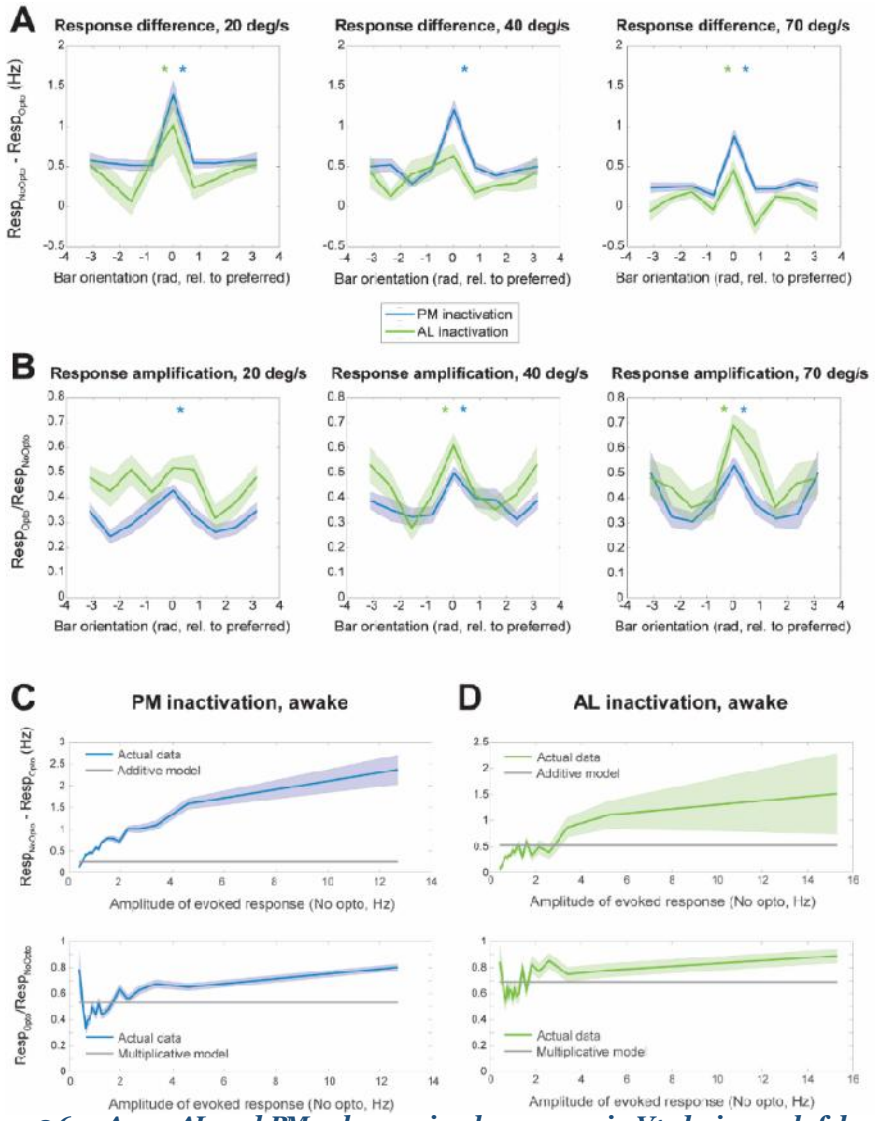
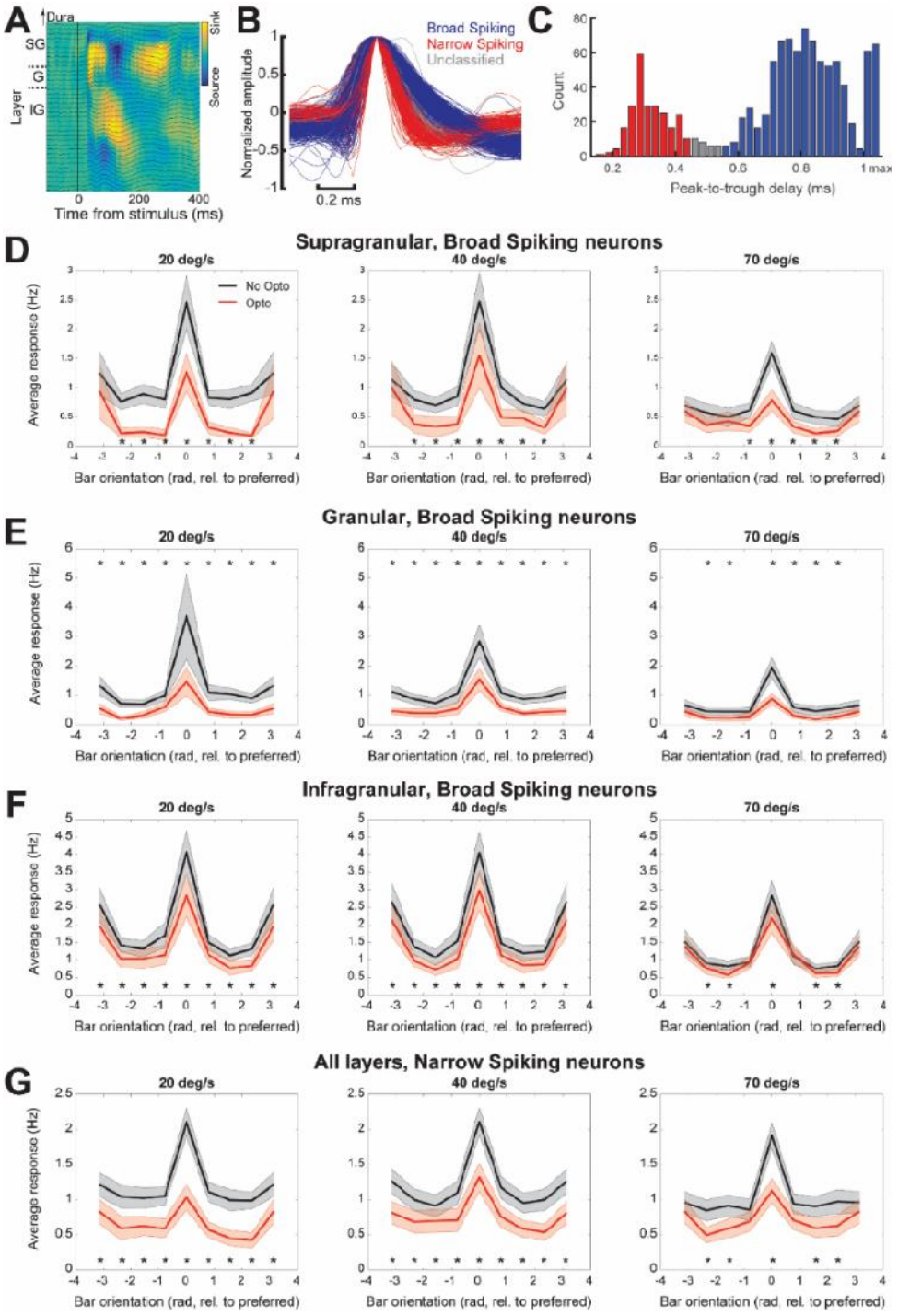


Figure 2.6 – Areas AL and PM enhance visual responses in V1 during wakefulness neither by additive nor multiplicative gain. A. *Difference in sensory-evoked responses of V1 neurons in the absence or presence of optogenetic inactivation of either area PM (blue lines) or AL (green lines) during awake recordings, separately for each orientation of a moving bar – respectively – at 20 deg/s (left), 40 deg/s (middle) or 70 deg/s (right). Curves with shading indicate mean ± SEM. The asterisks indicate if, separately for each bar speed and area (as indicated by color of asterisk), the difference in sensory evoked responses was the same for all orientations (null hypothesis, corresponding to a flat line in the plot), or not ($p<0.05$, one way Anova; a post-hoc Tukey test revealed a larger difference for the preferred orientation corresponding to 0 rad; for PM-20deg/s: $F=12.43$, $p=1.4x10^{-15}$; for PM-40deg/s: $F=15.02$, $p=4.0x10^{-19}$; for PM-70deg/s: $F=17.12$, $p=7.0x10^{-22}$; for AL-20deg/s: $F=2.54$, $p=0.0136$; for AL-40deg/s: $F=1.68$, $p=0.1102$; for AL-70deg/s: $F=4.66$, $p=4.2x10^{-5}$).* ***B.*** *Same as A, but now for the ratio between sensory-evoked responses of V1 neurons in the presence or absence of optogenetic inactivation of either area PM (blue lines) or area AL (green lines). In general, responses to the preferred orientation were reduced less by optogenetic*

inactivation of area PM or AL, compared to responses to non-preferred orientations ($p<0.05$, one-way Anova with post-hoc Tukey test; for PM-20deg/s: $F=3.59$, $p=0.0008$; for PM-40deg/s: $F=2.87$, $p=0.0056$; for PM-70deg/s: $F=3.01$, $p=0.0038$; for AL-20deg/s: $F=1.91$, $p=0.0645$; for AL-40deg/s: $F=3.39$, $p=0.0014$; for AL-70deg/s: $F=2.46$, $p=0.0168$). ***C.*** *Curves showing difference and amplification (top and bottom panel, respectively) of V1 responses following PM inactivation as a function of the amplitude of sensory-evoked responses under unperturbed conditions. Sensory-evoked responses (x-axis) were divided into 20 equipopulated bins and, for each bin, the mean ± SEM value of optogenetic-induced responses difference or amplification was computed (solid blue line and shading, respectively). The grey line indicated the response difference or amplification obtained when applying an additive (top panel) or a multiplicative (bottom panel) model – see Materials and Methods for details. Optogenetic inactivation of area PM modulates sensory-evoked V1 responses in a way that is different from both an additive or multiplicative model. In particular, the bottom plot shows how optogenetic inactivation of area PM more strongly suppressed low-amplitude compared to high-amplitude sensory-evoked responses ($p<2x10^{-16}$, Chi-square goodness of fit test against a uniform distribution).* ***D.*** *Same as E, but now for the inactivation of area AL. As for inactivation of PM, inactivation of AL modulates sensory-evoked V1 responses in a way that is different from both an additive or multiplicative model. Also, low-amplitude sensory-evoked responses are more strongly suppressed compared to high-amplitude sensory-evoked responses ($p<2x10^{-16}$, Chi-square goodness of fit test against a uniform distribution). The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*

evoked by bars moving along non-preferred orientations and directions. To investigate whether this amplitude-dependent modulation applies to all V1 neurons, independently from their response amplitude and orientation selectivity, we stratified V1 neurons based on either the amplitude of the sensory-evoked responses or the value of gOSI. Irrespective of whether we focused on the top or bottom 50% for the responses to the preferred orientation or the gOSI, we found that optogenetic inactivation of PM and AL invariably decreased responses to both preferred and non-preferred orientations (although the latter were more strongly impacted) and enhanced gOSI (Supplementary Fig. 2.12A-D). This suggests that PM and AL do not implement a selective suppression of weakly responsive or non-selective neurons, or specifically reduce (or even suppress) low-amplitude responses. Rather, inactivation of PM and AL similarly affects all V1 neurons and sensory-evoked responses, albeit in a way that non-linearly depends on response amplitude (Fig. 2.6C-D). Furthermore, we asked if the enhanced orientation and direction selectivity that we reported following inactivation of PM and AL might be amplified or even caused by the occurrence of few, spurious spikes occurring in neurons which would have otherwise been silenced. To answer this question, we focused on the subset of neurons showing gDSI=1 in the opto condition. These neurons maintained their direction preference (albeit with lower direction and orientation selectivity) in the no-opto condition (Supplementary Fig. 2.12E), which indicated that even these extreme values of gDSI are not an artefactual consequence of optogenetic inactivation.


A
Dura
SG
Layer
G
IG
Sink
Source
0
200
400
Time from stimulus (ms)
B
Broad Spiking
Narrow Spiking
Unclassified
Normalized amplitude
0.2 ms
C
Count
Peak-to-trough delay (ms)
1 max
D
Supragranular, Broad Spiking neurons
20 deg/s
40 deg/s
70 deg/s
No Opto
Opto
Average response (Hz)
Bar orientation (rad, rel. to preferred)
E
Granular, Broad Spiking neurons
F
Infragranular, Broad Spiking neurons
G
All layers, Narrow Spiking neurons

(figure on the previous page) Figure 2.7 – Optogenetic inactivation of area AL and PM during wakefulness similarly affects distinct V1 neuronal subpopulations. A. *Current source density profile of average response to checkerboard stimulation across layers of V1. The bottom of the earliest sink after checkerboard onset was used to demarcate the boundary between L4 and L5 – as in* (Schnabel et al., 2018). *We used this in combination with registered depth of penetration of the silicon probe relative to the cortical surface, to align electrode depth across recordings sessions. Black traces show the local field potential traces of each channel along the electrode tract.* ***B.*** *Normalized waveforms for each individually recorded V1 neuron, averaged over all recorded action potentials and colored by their classification based on peak-to-trough delay (blue: broad spiking neuron; red: narrow spiking neuron; grey: undetermined).* ***C.*** *Histogram of peak-to-trough delay for the three classes. Bars at maximum indicate neurons whose trough extended beyond the sampled time around the action potential (2 ms).* ***D.*** *Tuning curves of average responses of broad spiking V1 neurons located in supragranular layers of V1 during awake recordings to bars moving at different orientation (aligned by the preferred orientation, which is set at 0 deg) in the absence or presence of optogenetic inactivation (black and red curves, respectively). Recordings performed during the inactivation of PM and AL were pooled together. Curves with shading indicate mean ± SEM. Asterisks indicate significant differences between responses to bars moving at a given orientation in the absence or presence of optogenetic inactivation ($p<0.05$, paired t-test, FDR-corrected). From left to right: responses to bars moving at 20, 40 and 70 deg/s. The number of neurons included in this analysis was 34 (20 deg/s), 34 (40 deg/s) and 29 (70 deg/s).* ***E.*** *Same as D, but now for broad spiking V1 neurons in granular layers. The number of neurons included in this analysis was 28 (20 deg/s), 30 (40 deg/s) and 28 (70 deg/s).* ***F.*** *Same as D, but now for broad spiking V1 neurons in infragranular layers. The number of neurons included in this analysis was 161 (20 deg/s), 159 (40 deg/s) and 139 (70 deg/s).* ***G.*** *Same as D, but now for narrow spiking V1 neurons. The number of neurons included in this analysis was 67 (20 deg/s), 68 (40 deg/s) and 56 (70 deg/s).*

To further explore the mechanism underlying this form of gain modulation, and in particular whether AL and PM targeted specific neuronal subpopulations in V1, we subdivided neurons based on whether their action potential waveform was broad or narrow, i.e. corresponding to a putative pyramidal or fast-spiking interneuron, respectively (Olcese et al., 2013, 2016; Vinck et al., 2015b), and based on whether neurons were located in supragranular, granular or infragranular layers (see Fig. 2.7A-C and Materials and Methods). However, no difference was observed between putative excitatory and inhibitory neurons, nor between putative excitatory neurons residing in different cortical layers as concerns changes in visual responses following inactivation of AL and PM (Fig. 2.7D-G). Therefore, the modulation provided by AL and PM onto V1 seems to similarly affect all major neuronal components of V1.

Discussion

Areas AL and PM impact mostly on weak sensory-evoked responses in V1

Inactivation of either area AL or PM similarly reduced sensory-evoked responses in V1 (Fig. 2.2). Therefore, the main role of AL and PM, in our paradigm, is to enhance sensory-evoked responses in V1, in particular weaker ones such as those to non-preferred stimuli (Fig. 2.6) and those occurring in the early phase of sensory-evoked responses (Fig. 2.3). It was previously shown that inactivating HVAs (Pafundo et al., 2016) or feedback projections from HVAs onto V1 (Huh et al., 2018) decrease sensory-evoked responses in V1. However, in previous studies the effects were either limited to responses to the preferred orientation (Pafundo et al., 2016), were specific for the receptive field center and not the surround (Nurminen et al., 2018), or were functionally specific (i.e. related to the tuning properties

of individual HVAs). For instance, Huh et al. (2018) reported that inactivation of feedback projections from AL or PM onto V1 specifically reduced responsiveness of V1 neurons tuned to spatial frequencies similar to those of AL or PM. In contrast, we observed a generalized decrease in V1 responses and receptive field size that was independent of the speed tuning of either AL, PM, or individual V1 neurons. This functionally aspecific effect (at least for what pertains speed tuning for moving bars) is even more unexpected when considering that anatomical projections from HVAs onto V1 are also functionally specific (Marques et al., 2018) and target patches in layer 1 of V1 (D'Souza et al., 2019; Ji et al., 2015) based on factors such as orientation/direction tuning and speed preference. How are these two apparently discordant effects compatible?

First, most previous studies focused on the lateromedial (LM) secondary visual area (Marques et al., 2018; Pafundo et al., 2016). LM is thought to be part of the mouse ventral stream, while both AL and PM are attributed to the dorsal stream (Wang et al., 2012). Thus, the roles of area LM versus AL and PM in modulating activity in V1 might be different, also in view of the specific functions of LM in higher order visual processing (Khastkhodaei et al., 2016; Matteucci et al., 2019; Pak et al., 2020; Tafazoli et al., 2017). Inactivation of LM primarily affects the superficial layers of V1 (Pafundo et al., 2016); conversely, in our study inactivation of AL or PM similarly affected all the major subpopulations of V1. Furthermore, none of the prior studies used moving bars, but drifting gratings. Gratings are commonly displayed over a larger field of view compared to bars (which, in our case, were only 3 deg wide stimuli). Thus, moving gratings simultaneously evoke activity in a larger population of neurons compared to bars. Therefore, at the population level, moving bars elicit an overall instantaneous weaker activity compared to gratings. The aspecific modulatory effect we report primarily affects weak responses to visual stimuli, but it is possible that it might occur jointly with a functionally specific form of modulation (i.e. dependent on the spatial and temporal tuning properties of HVAs and V1 neurons), such as that described for instance in Huh et al. (2018). Indeed, when we performed preliminary experiments with drifting gratings, we found results in line with Huh et al. (2018). Nevertheless, it is striking that we did not observe any functionally-specific effect with bars moving at different speeds, given that speed tuning of AL and PM are among the most different among HVAs (Andermann et al., 2011; Marshel et al., 2011). Another difference between our study and Huh et al. (2018) is that we completely inactivated AL and PM, and not just the neurons in these areas projecting back to V1. It may be the case that inactivation of a network node, and not just of feedback-projecting neurons, may have a broader, less specific effect, which may be due to a combination of direct and indirect pathways (i.e., direct feedback projections from AL and PM to V1, as well as pathways involving other cortical regions – in particular other HVAs – as well as cortico-thalamic loops).

An additional contrast between our and previous studies lies in the minor differences that we observed between the effect of inactivating AL and PM during wakefulness or isoflurane anesthesia. In spite of differences in speed preference between wakefulness and anesthesia (Fig. 2.1J), the effect of optogenetic inactivation of HVAs on V1 responses was very similar between the two brain states (see for instance Fig. 2.2 and Supplementary Fig. 2.5). Conversely, other studies reported that inactivation of HVAs more strongly affects V1 during wakefulness than anesthesia (Keller et al., 2020b; Vangeneugden et al., 2019). This weaker effect of top-down modulation under anesthesia is in line with results from human subjects during loss of consciousness (Boly et al., 2011; Sikkens et al., 2019). A likely

explanation for these different results lies again in the different type of stimulus that we used, which evokes less powerful activity changes in visual cortices compared to gratings.

Enhanced sensory-evoked responses in V1 come at the expense of orientation and direction selectivity

The enhancement of sensory-evoked responses that AL and PM induce in V1 comes, strikingly, at the expense of orientation and direction selectivity, both at the single-neuron and population level. This result is particularly surprising given the fact the previous studies generally showed that HVAs enhance visual processing in V1. One study in particular (Pafundo et al., 2016) reported that inactivation of area LM decreased V1 responses to gratings moving along single neurons' preferred orientation, and consequently also reduced orientation and direction selectivity of V1 neurons. In contrast, we observed a marked enhancement of orientation and direction tuning when AL and PM were inactivated (see the previous subsection for a discussion or possible reasons underlying the different results we observed). This finding suggests that the weaker orientation and direction tuning which is present in V1 with functionally intact AL and PM is likely sufficient to enable a proper processing of visual stimuli – although different results might have been obtained with other types of visual stimuli, such as moving gratings (e.g. (Jin and Glickfeld, 2020)). Therefore, the visual system might operate in a regime that balances the processing of stimulus features such as orientation and direction with the ability to process stimuli which are smaller and (at least at the single-neuron level) less salient.

A second interpretation is that HVAs such as AL and PM might provide contextual, predictive representations to V1 via recurrent projections. Following a predictive processing framework (Friston, 2005; Pennartz et al., 2019; Rao and Ballard, 1999), higher-order feedback may modulate V1 based on spatiotemporal predictions of sensory input, for example modulating neuronal activity of V1 neurons whose receptive field lie along the expected trajectory of a moving object (Marques et al., 2018). Our findings that HVA's specifically modulate V1 responses during the rising phase of the response (as the moving bar enters their receptive field) is in line with this interpretation and suggests that HVA preferentially play a role in shaping the response of V1 neurons as sensory input is expected to hit their receptive field. Such HVA-mediated prediction-related enhancement in responses may be, from a functional point of view, more relevant than a further increase in orientation/direction selectivity. Moreover, this framework may explain why our results differ considerably from previous studies that used moving gratings, because the latter type of stimulus conveys a much higher spatial predictability across the visual field than an isolated moving bar.

Of relevance, the effects of AL and PM inactivation were similar on single-neuron orientation and direction tuning, but different in terms of population decoding: inactivating AL improved population decoding of stimulus direction more than inactivation of PM did (Fig. 2.5). Recent studies also identified different functions of AL and PM in orientation discrimination and spatial integration, with PM showing larger receptive fields than AL (Murgas et al., 2020) and no involvement (in contrast with AL) in orientation discrimination (Jin and Glickfeld, 2020).

Finally, it is worthwhile to highlight that the increased orientation and direction selectivity that we reported following inactivation to AL and PM are unlikely to be a consequence of

an “iceberg” effect, in which most responses are silenced with the exception of the strongest ones. Indeed, we observed a reduction in responses (but not a complete suppression) in all neuronal cohorts we analyzed, irrespective of response amplitude or orientation/direction selectivity (Supplementary Fig. 2.12).

Higher order visual areas enhance stimulus responsiveness in V1

The classical framework to interpret visual processing in the neocortex follows a hierarchical approach, in which each subsequent processing stage is tasked with processing more complex stimulus features (Felleman and Van Essen, 1991; Riesenhuber and Poggio, 1999). HVA properties seem to support this view, as some higher-order visual processing is either directly dependent upon, or facilitated by them (Khastkhodaei et al., 2016; Matteucci et al., 2019; Pak et al., 2020; Tafazoli et al., 2017). Nevertheless, recent studies have shown that some key features of early visual processing, at the stage of V1, are enabled by virtue of top-down modulation originating in HVAs (Keller et al., 2020b; Vangeneugden et al., 2019). Our study supports this notion, by indicating that, beyond being involved in the development of response features such as surround suppression (Vangeneugden et al., 2019), complex receptive fields (Keller et al., 2020b) and higher-level representations (Pak et al., 2020), feedback information from HVAs also contributes to basic properties of V1 such as responses to oriented bars. Our experiments showed that AL and PM enhance weak sensory-evoked V1 responses more than strong ones. This may be explained by an added level of non-specific background excitation that HVAs could provide to V1 neurons. Such additional excitation might modify the supposedly sigmoid input-output transfer function of V1 neurons in a way that more strongly amplifies weak inputs compared to strong ones. Other mechanisms, such as non-linearities intrinsic to the generation of action potentials, cannot however be excluded.

By enhancing stimulus responsiveness, in particular to non-preferred and small, unexpected visual features, just entering single neurons’ receptive fields, HVAs such as AL and PM might play a role akin to that fulfilled by inverse effectiveness in the context of multisensory cue integration (Meijer et al., 2019; Stein and Stanford, 2008). Specifically, HVAs might enhance, in particular at the single-neuron level, the signal-to-noise ratio of sensory evoked responses to stimuli that would not otherwise induce large responses (for instance non-preferred directions of movement, or stimuli entering the receptive field). This may provide a behavioral advantage by enabling to more reliably process small, barely noticeable visual stimuli.

Conclusions

Higher order visual areas are key elements of the cortical network of visual processing, as they not only further analyse visual information coming from V1, but also modulate the activity of V1 itself. Here we showed how two HVAs with different response properties similarly enhance responses of V1, especially weak and unexpected ones (such as responses to non-preferred directions, or responses to bars entering a neuron’s receptive field). At the population level, AL and PM activity makes it easier for V1 to respond to moving bars, but at the same time more difficult to decode their precise orientation and direction. Areas AL and PM therefore provide a major contribution to sculpting of V1 responses to simple visual objects: they effectively contribute to generating stronger and less sparse responses which, in turn, might make sensory-evoked responses to small, local and possibly unexpected stimuli such as moving bars more robust and ultimately more reliable.

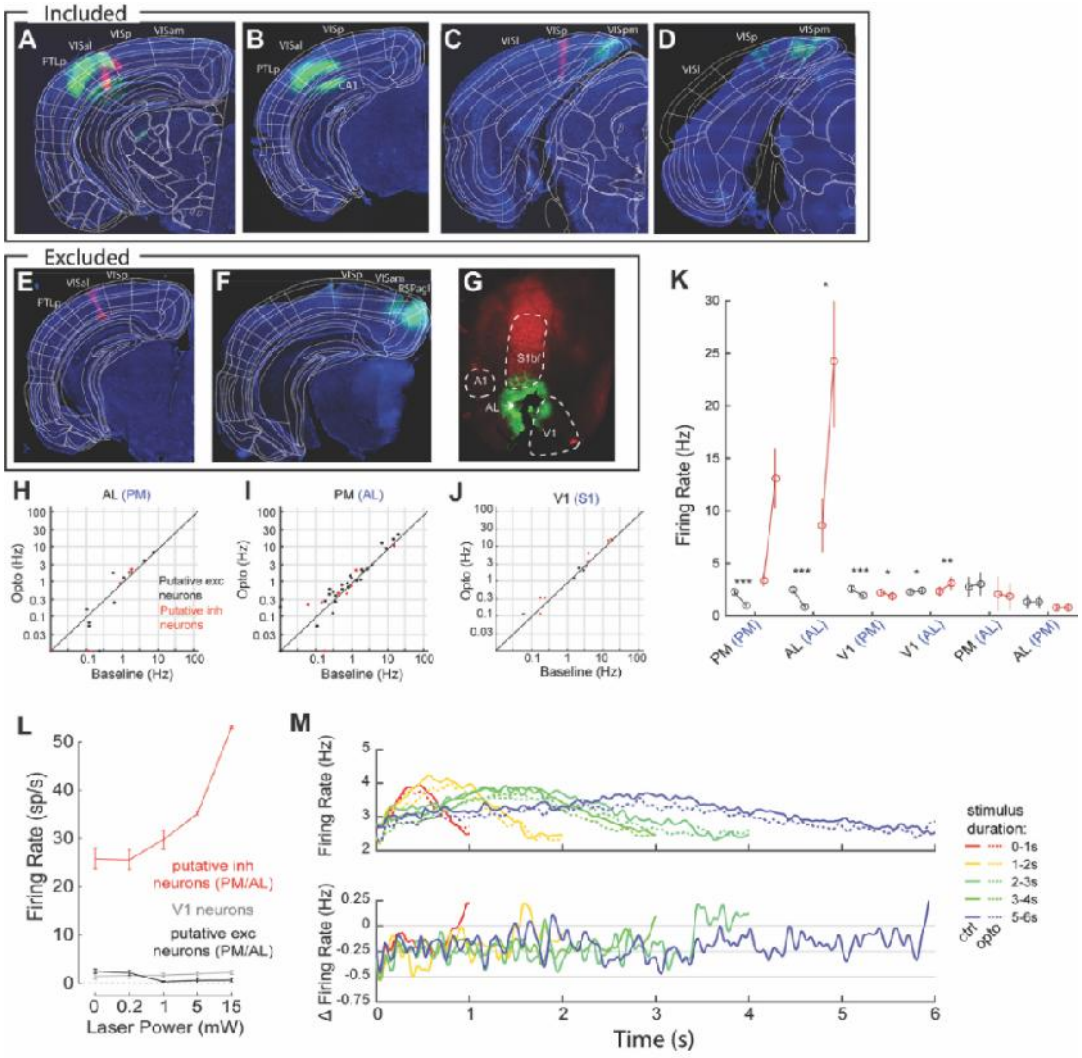
Funding

This work was supported by the European Union's Horizon 2020 Framework Program for Research and Innovation (Specific Grant Agreement 720270 - Human Brain Project SGA1 - to C.M.A.P., Grant Agreement 785907 - Human Brain Project SGA2 - and 945539 - Human Brain Project SGA3 - to C.M.A.P. and U.O.) and by FLAG-ERA - co-financed by the Netherlands Organization for Scientific Research (Joint Transnational Call 2015, project CANON to U.O. and Joint Transnational Call 2019, project DOMINO to U.O).

Acknowledgements

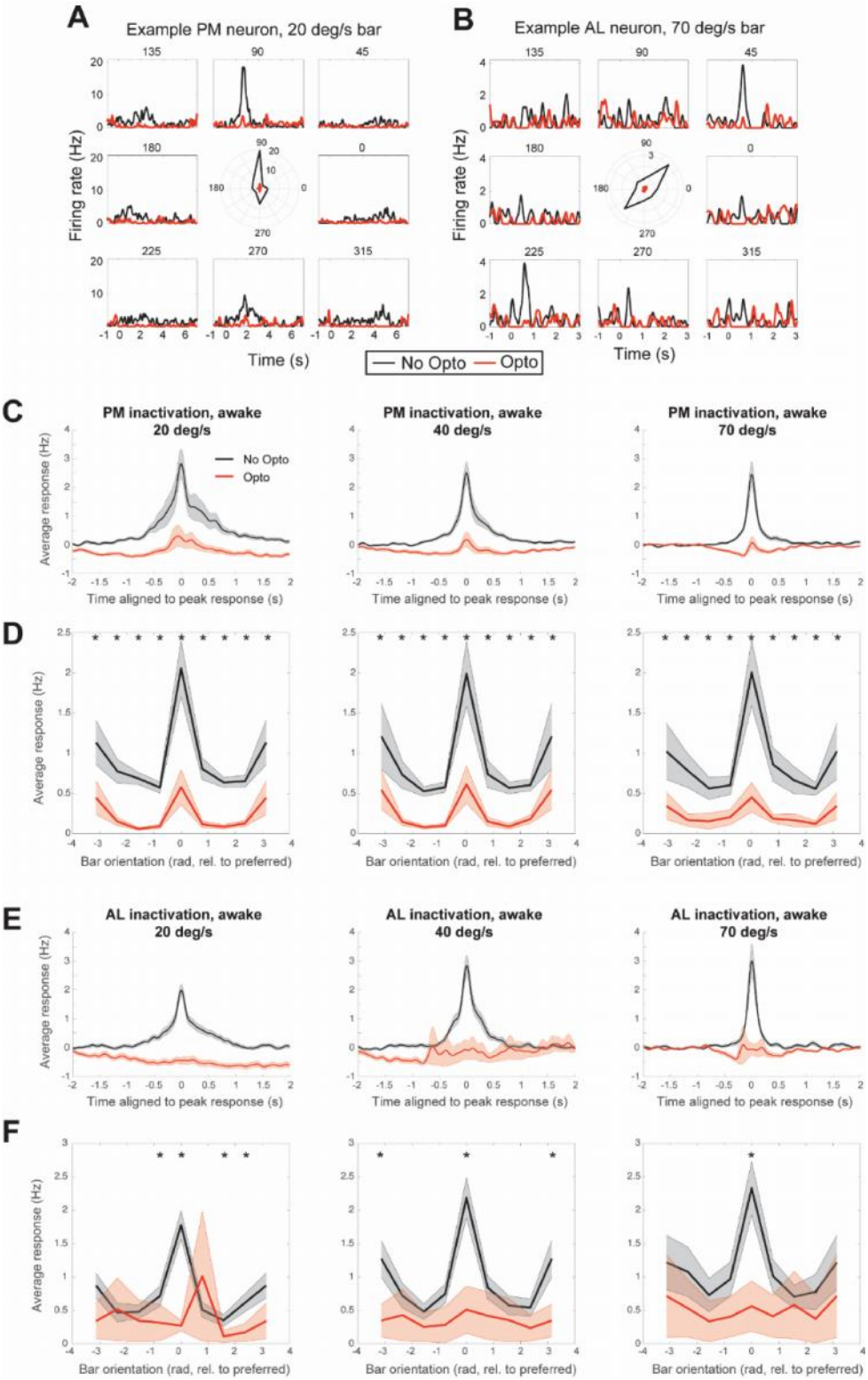
The authors would like to thank Laura Bavelaar for her support in the initial phases of this project, and the anonymous reviewers for their insightful comments, which have greatly improved our manuscript.

Supplementary Figures

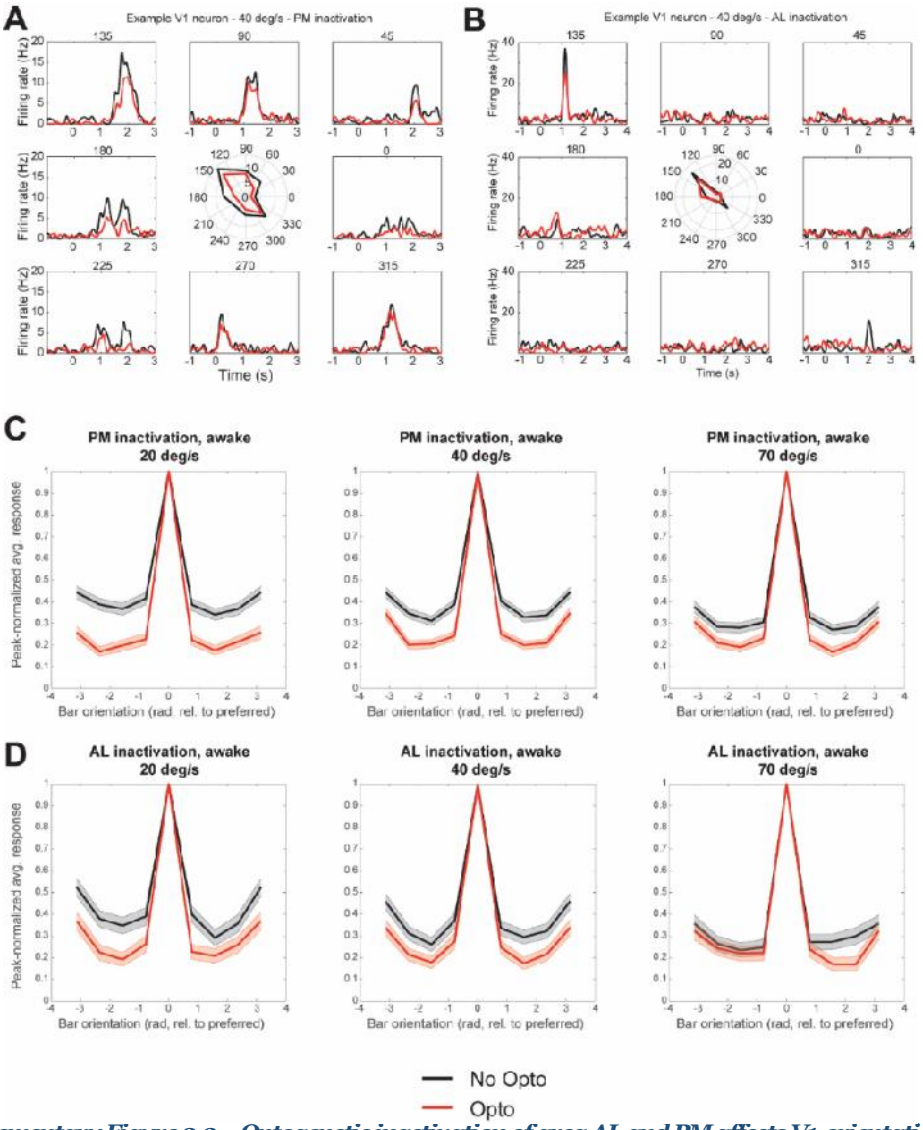


Supplementary Figure 2.1 – Optogenetic inactivation of area AL and PM silences responses to moving stimuli in the illuminated area. A. *Coronal section showing ChR2-conjugated GFP expression in area AL (green) and the location of a laminar probe positioned in AL and stained with DiI (red). Area nomenclature follows that of the Allen Brain Atlas: VISal: area AL; VISp: area V1; VISam: anteromedial (AM) secondary visual cortex; PTLp: posterior parietal cortex.* ***B.*** *Same as A, for an additional example animal with target area AL.* ***C.*** *Coronal section showing ChR2-conjugated GFP expression in area PM (green) and the location of a laminar probe positioned in V1 and stained with DiI (red). VISpm: area PM; VISl: lateral secondary visual cortex.* ***D.*** *Same as C, for an additional example animal with target area PM. A-D show examples of different animals included in the dataset based on localized expression.* ***E.*** *Same as A, but for an excluded animal where no expression was observed in target area AL.* ***F.*** *Same as C, but for an excluded animal with target area PM where strong off-target expression was observed in retrosplenial cortex. RSPagl: Retrosplenial cortex, agranular part.* ***G.*** *Same as Fig. 2.1C, but for an excluded animal with target*

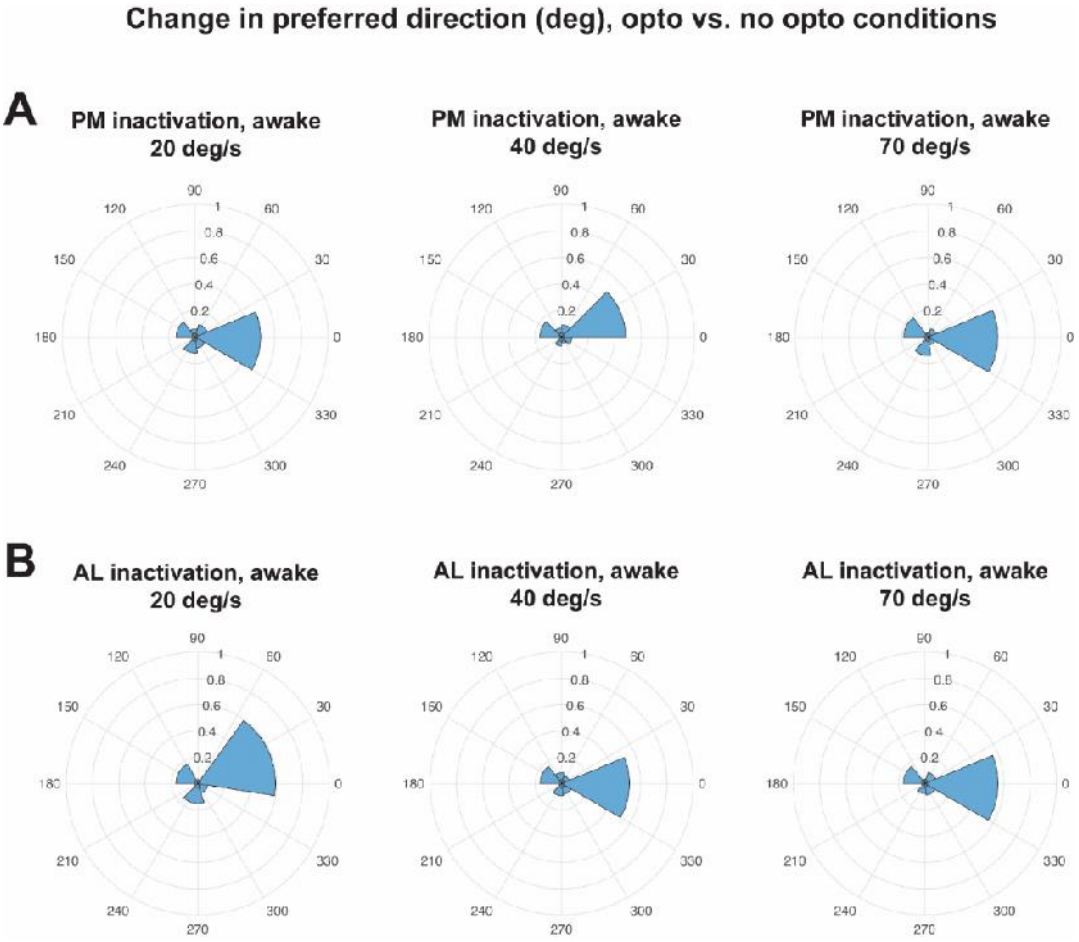
area AL with expression in S1 bf, other HVA's, and V1. S1 bf: Barrel field of primary somatosensory cortex. ***H.*** *Same as Fig. 2.1 G,H, but now for the spontaneous activity of neurons recorded in AL during optogenetic stimulation of PM. For both cell types, spontaneous activity was not significantly affected (putative excitatory neurons: n=12 neurons, 2.24 Hz to 2.23 Hz, p=0.700; putative inhibitory neurons: n=6 neurons, 1.53 Hz to 1.61 Hz, p=0.625, Wilcoxon Signed Rank test). This further highlights that the inactivation was local to the photostimulated area.* ***I.*** *Same as H, but for neurons recorded from PM during optogenetic stimulation of AL. For both cell types, spontaneous activity was not significantly affected (putative excitatory neurons: n=26 neurons, 2.9 Hz to 3.1 Hz, p=0.99; putative inhibitory neurons: n=8 neurons, 2.1 Hz to 1.9 Hz, p=0.94, Wilcoxon Signed Rank test).* ***F.*** *Same as Fig. 2.1 G,H, but now for the spontaneous activity of neurons recorded in V1 during laser illumination of uninfected control area S1. For both cell types, spontaneous activity was not significantly affected (putative excitatory neurons: n=7 neurons, 1.46 Hz to 1.19 Hz, p=0.156, putative inhibitory neurons: n=8 neurons: 4.90 Hz to 5.08 Hz, p=0.688, Wilcoxon Signed Rank test). This excludes the possibility that direct activation of the retina affected activity in visual cortex.* ***K.*** *Quantification of all scatterplots show in Main Fig. 2.1 G,H and in panels H and I of this supplementary figure. Each condition shows spontaneous activity during baseline (left) and photostimulation (right) for neurons recorded in a given area (black font) with optogenetic stimulation in another area (blue font in brackets). For values and result of significance testing see legend of associated scatterplots. Dot and errorbar denote mean ± SEM across neurons (black: putative excitatory neurons; red: putative inhibitory neurons). The slight increase in activity of putative inhibitory neurons in V1 upon AL illumination is unlikely to be the cause of direct optogenetic activation as it is simultaneously associated with an increase in firing in excitatory neurons.* ***L.*** *Spontaneous activity during photostimulation at increasing photostimulation power (dose-response curve) for putative inhibitory and excitatory neurons in the photostimulated area (PM and AL neurons combined, n=31 neurons; black: putative excitatory neurons; red: putative inhibitory neurons), or V1 neurons (n=36 neurons, grey). Putative inhibitory neurons in PM/AL were increasingly recruited by higher laser powers, while putative excitatory neurons were increasingly suppressed. V1 neurons were minimally affected, independently from laser power.* ***M.*** *To test whether photostimulation had non-stationary effects, e.g. due to heating of the illuminated cortical tissue, we investigated the effect of photostimulation on V1 activity as a function of stimulus duration (and consequently also photoillumination duration). Trials (all orientations) were binned based on the duration of stimulus presentation (bars moving at 70 deg/s had on average a stimulus durations shorter than 2 seconds, and bars moving at 20 deg/s more than 3 seconds). The upper panel shows the firing rate for control and photostimulated trials. The lower panel show the change in firing rate by photostimulation (opto - control) indicating that the effect of PM or AL inactivation on V1 firing rates, when averaged over all stimulus orientations and neurons, was independent of duration of photostimulation, i.e. the change in firing rate remains constant over time.*


A
Example PM neuron, 20 deg/s bar
135
90
45
180
0
225
270
315
Firing rate (Hz)
Time (s)
B
Example AL neuron, 70 deg/s bar
Firing rate (Hz)
Time (s)
No Opto
Opto
C
PM inactivation, awake
20 deg/s
40 deg/s
70 deg/s
Average response (Hz)
Time aligned to peak response (s)
D
Bar orientation (rad, rel. to preferred)
E
AL inactivation, awake
20 deg/s
40 deg/s
70 deg/s
F

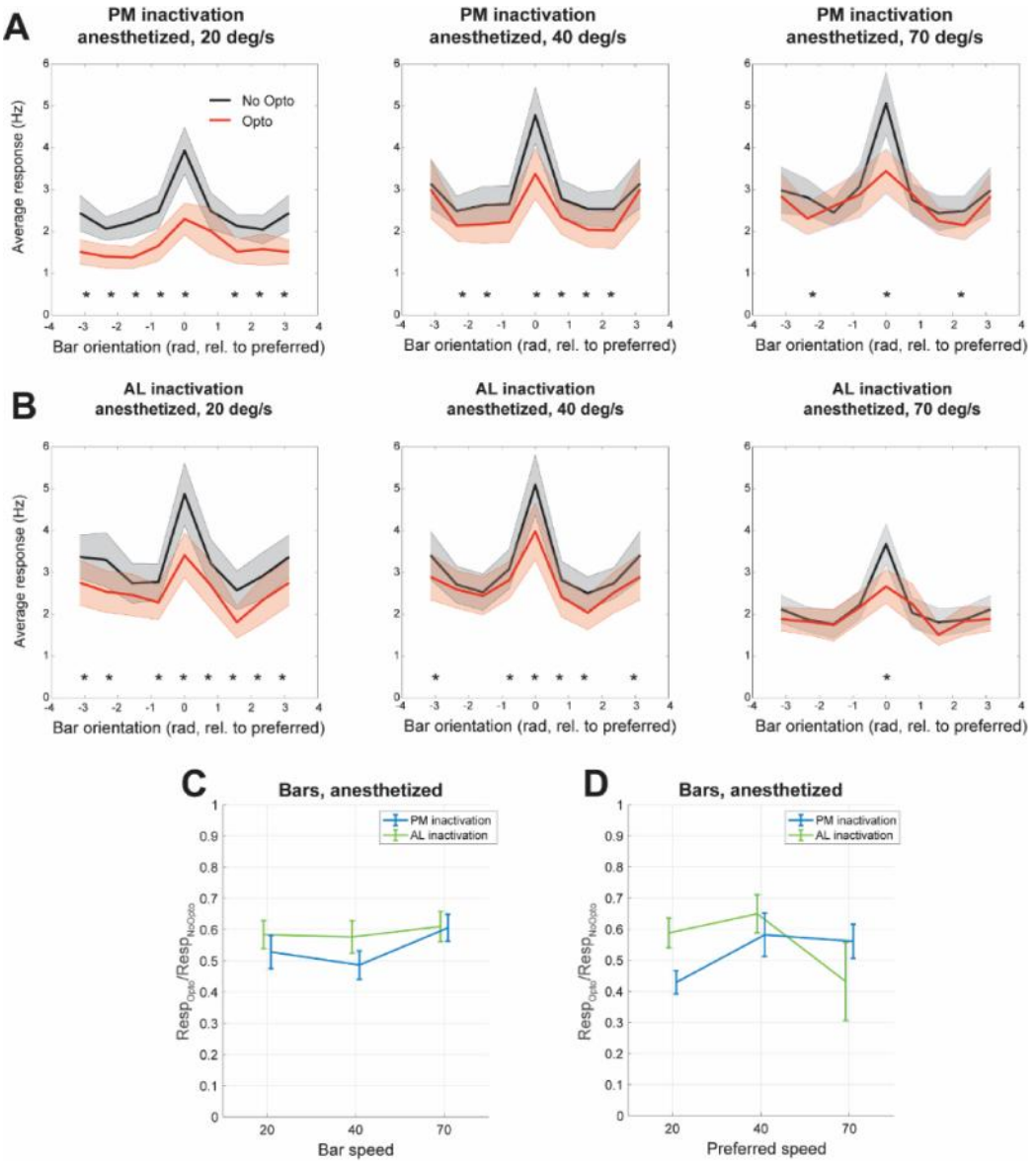
Supplementary Figure 2.2 – Optogenetic inactivation of area AL and PM silences responses to moving stimuli in the illuminated area. A. *Same as Fig. 2.1K, but now for a neuron located in area PM and responding to a bar moving at 20 deg/s during wakefulness. Black: control; red: inactivation of area PM.* ***B.*** *Same as A, but now for a neuron located in area AL and responding to a bar moving at 70 deg/s during wakefulness. Note how optogenetic stimulation suppresses the responses of neurons located in PM and AL (panels A and B, respectively), but only reduces the responses of the neuron located in V1 (Fig. 2.1J).* ***C.*** *PSTHs (aligned to peak latency) averaged over all putative excitatory PM neurons responding to visual stimuli during awake recordings in the absence or presence of optogenetic inactivation of area PM itself (black and red curves, respectively). Curves with shading indicate mean ± SEM. From left to right: responses to bars moving at 20, 40 and 70 deg/s.* ***D.*** *Tuning curves of average responses of putative excitatory PM neurons during awake recordings to bars moving at different orientation (aligned by the preferred orientation, which is set at 0 rad) in the absence or presence of optogenetic inactivation of area PM itself (black and red curves, respectively). Curves with shading indicate mean ± SEM. Asterisks indicate significant differences between responses to bars moving at a given orientation in the absence or presence of PM inactivation ($p<0.05$, paired t-test, FDR-corrected). From left to right: responses to bars moving at 20 (127 neurons), 40 (127 neurons) and 70 deg/s (102 neurons).* ***E.*** *Same as C, but now for putative excitatory neurons in AL and optogenetic inactivation of area AL itself.* ***F.*** *Same as D, but now for putative excitatory neurons in AL and optogenetic inactivation of area AL itself. From left to right: responses to bars moving at 20 (50 neurons), 40 (55 neurons) and 70 deg/s (44 neurons). Notice how optogenetic inactivation reduces visually-evoked responses in PM and AL much more strongly than in V1 (cf. Fig. 2.2).*



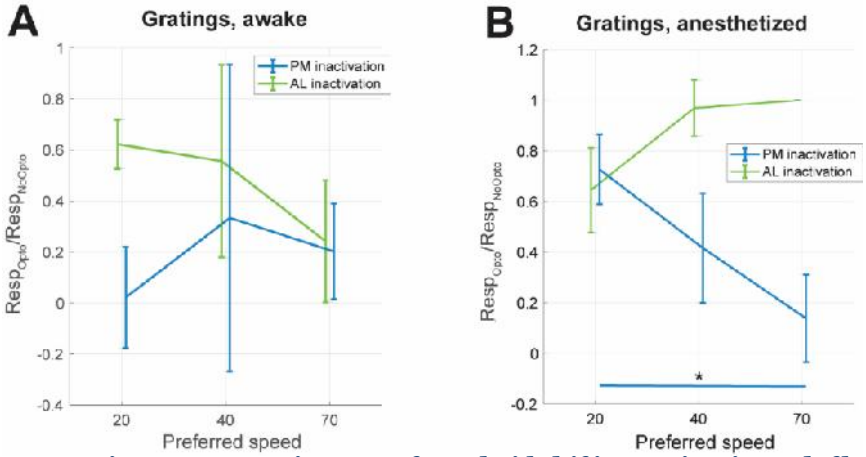
Supplementary Figure 2.3 – Optogenetic inactivation of area AL and PM affects V1 orientation and direction tuned responses. A. *Same as Fig. 2.1K, but for a neuron located in area PM and responding to a bar moving at 40 deg/s during wakefulness. Black: control; red: inactivation of area PM.* ***B.*** *Same as A, but now for a neuron located in area AL and responding to a bar moving at 40 deg/s during wakefulness. Note how optogenetic stimulation preserves the preferred orientation.* ***C.*** *Same as Fig. 2.2B, with responses normalized – individually for each recorded neuron and separately for No-Opto and Opto conditions – to the response to the preferred direction. This panel shows that, following optogenetic inactivation of area PM, neurons in V1 become more strongly tuned to both orientation and direction (notice the relatively weaker responses to non-preferred directions in opto compared to no-opto conditions).* ***D.*** *Same as panel C, but for optogenetic inactivation of area AL. The number of neurons included in the analyses shown in panels C and D is provided in Table 2.1.*



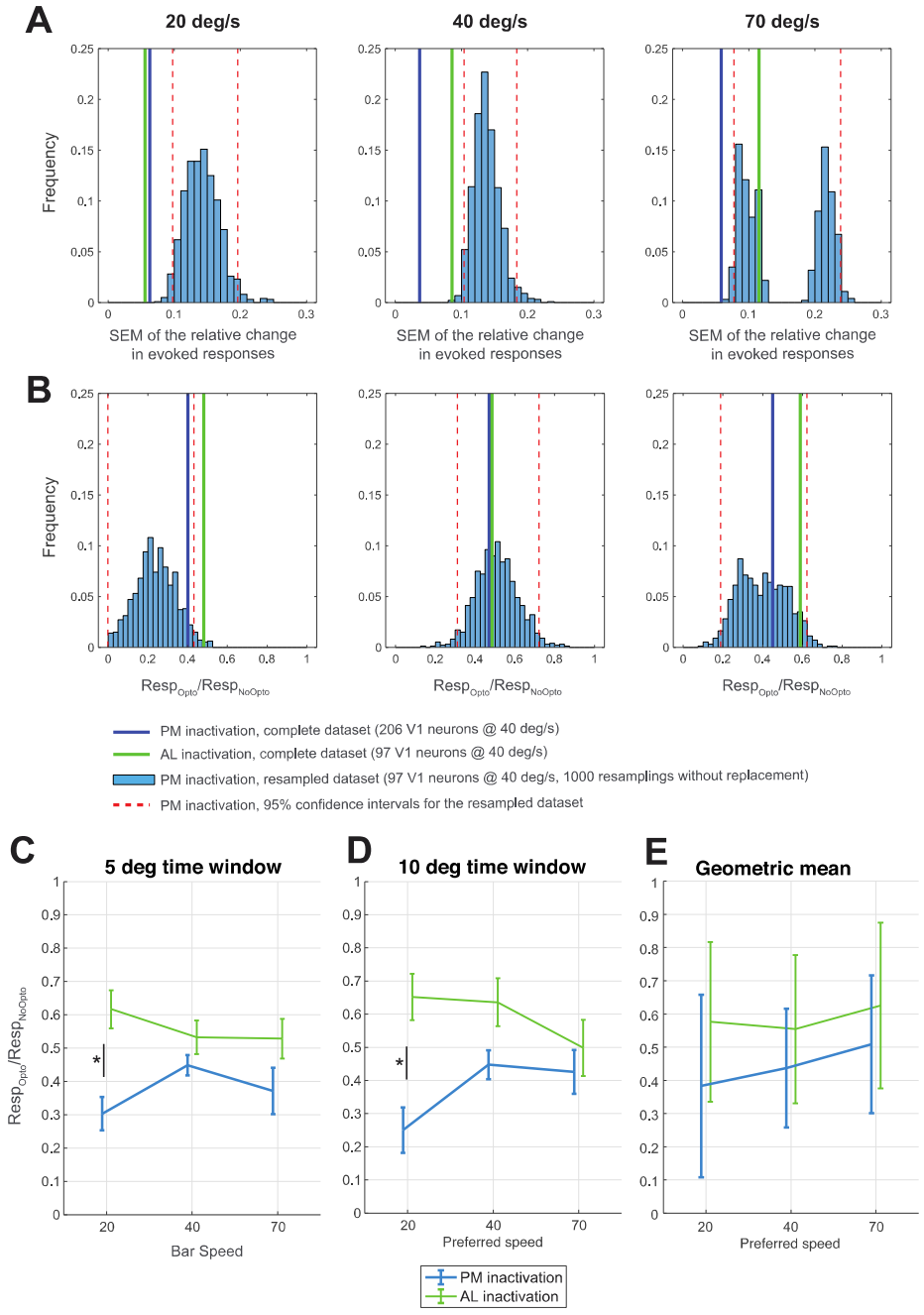
Supplementary Figure 2.4 – Optogenetic inactivation of area AL and PM does not modify the direction preferred by V1 neurons. A. *Polar histograms showing the distribution of the differences in preferred direction between No-Opto and Opto conditions following PM inactivation, as a function of bar speed (from left to right: 20, 40 and 70 deg/s). No significant difference was observed (one-sample test for mean angle).* ***B.*** *Same as A, but for the inactivation of area AL. Also in this case, no significant difference was observed. The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*



Supplementary Figure 2.5 – Optogenetic inactivation of area AL and PM under anesthesia depresses sensory-evoked responses in V1. A. *Same as Fig. 2.2B, but now for recordings performed under anesthesia.* ***B.*** *Same as Fig. 2.2D, but now for recordings performed under anesthesia.* ***C.*** *Same as Fig. 2.2E, but now for recordings performed under anesthesia.* ***D.*** *Same as Fig. 2.2F, but now for recordings performed under anesthesia. The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*

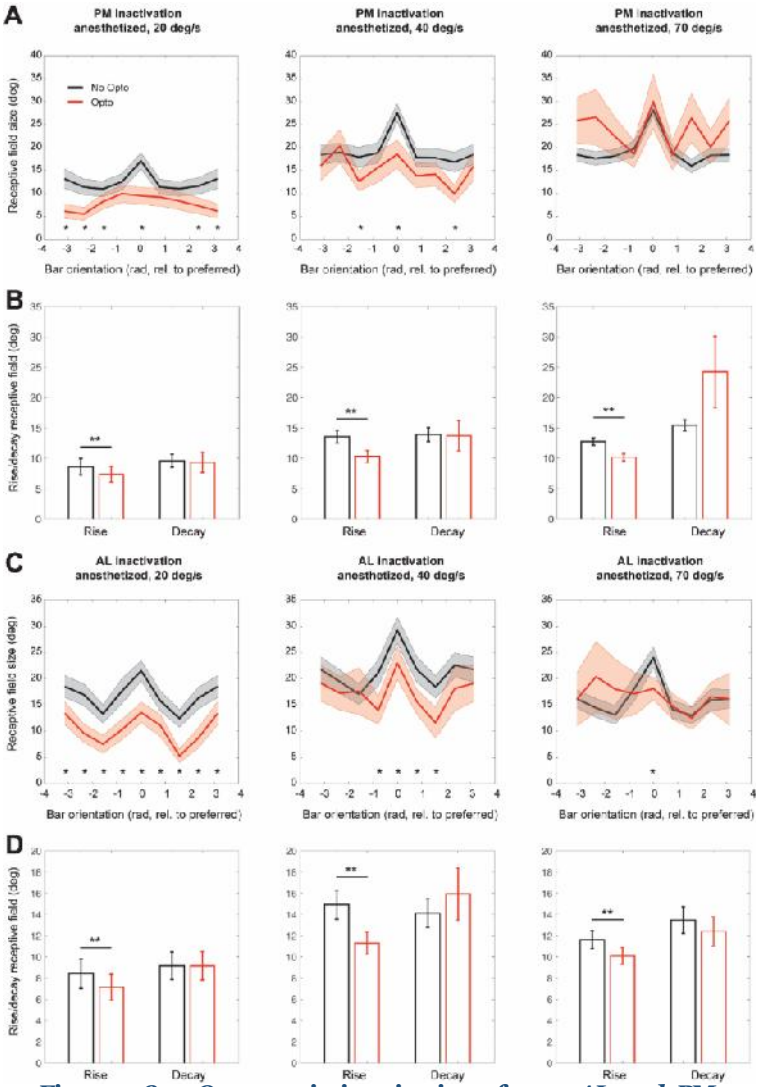


Supplementary Figure 2.6 – Experiments performed with drifting gratings instead of bars confirm the previously reported functionally-specific effect of inactivating AL and PM. A. *Same as Fig. 2.2F, but now for responses of V1 neurons to moving gratings during awake recordings. We included in this analysis 9 V1 neurons for PM inactivation and 70 for AL inactivation.* ***B.*** *Same as A, but now for recordings performed under anesthesia. Significant differences were observed for PM inactivation, between 20 and 70 deg/s (p=0.049, 2-way Anova with post-hoc Tukey test: significant main effect for inactivated area (F=8.34, p=0.0067)). We included in this analysis 27 V1 neurons for PM inactivation and 11 for AL inactivation.*

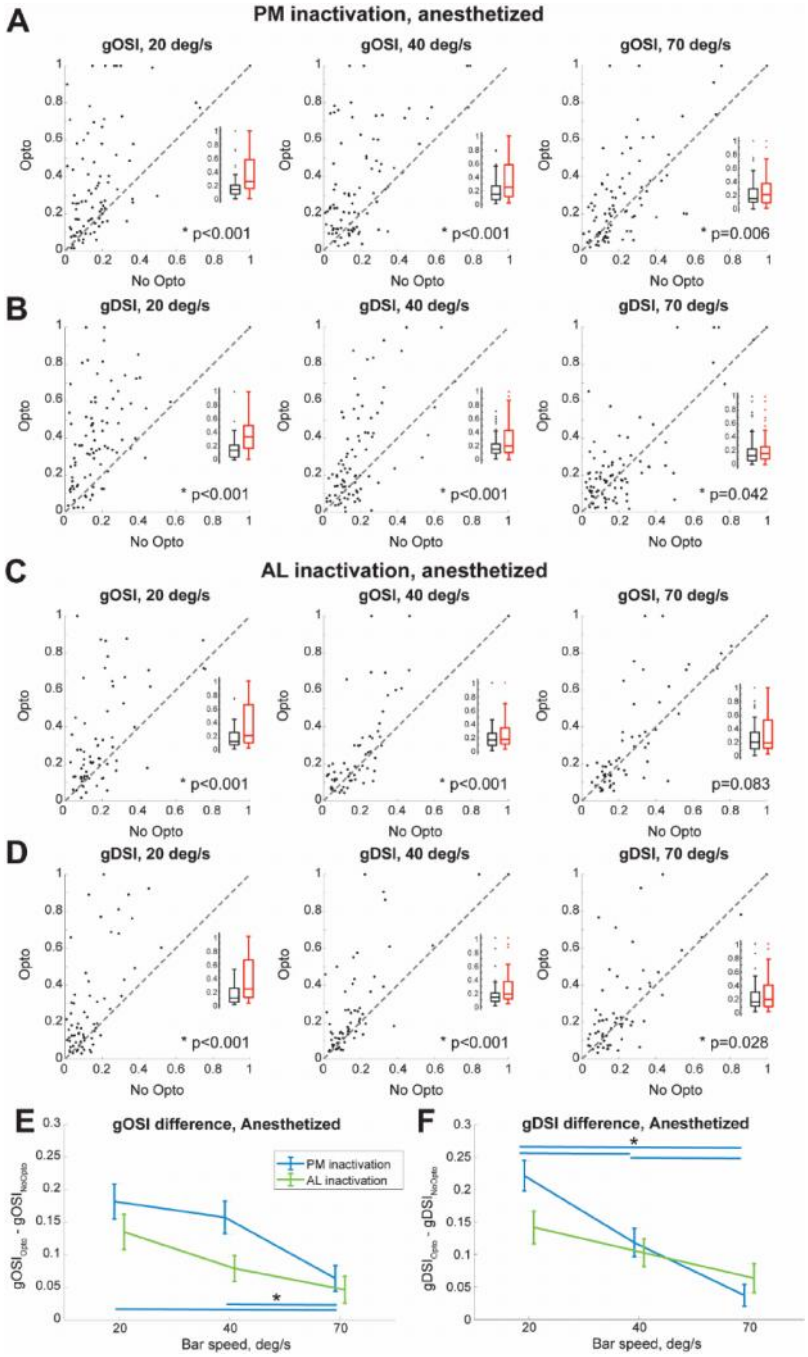


Supplementary Figure 2.7 – The effect of optogenetic inactivation of PM and AL on V1 responses is robust with respect to the number of recorded neurons and used metrics. A. *Histogram showing the distribution of SEM values for the relative change in evoked responses (amplitude modulation, see also Fig. 2.2E) computed over a resampled dataset of V1 neurons recorded in sessions in which PM*

was inactivated. In detail, resampling was performed 1000 times, and each time a number of neurons was sampled (without replacement) which corresponded to the number of V1 neurons recorded in sessions in which AL was inactivated (lower that neurons recorded for PM inactivation). Red vertical lines show the 95% confidence intervals for the SEM values computed on the resampled dataset. The blue and green vertical lines indicate, the SEM values obtained for the complete dataset for PM and AL inactivation, respectively. From left to right: awake recordings with bars moving at 20, 40 and 70 deg/s. ***B.*** *Same as A, but for the average value of the relative change in evoked responses following optogenetic modulation.* ***C.*** *Same as Fig. 2.2E, but with sensory-evoked responses computed over a time window whose duration was not fixed, but variable as a function of bar speed, in a way that such window always corresponded to the time a bar took to travel over a 5 deg window centered around peak response (individually for each neuron). P=0.0002, 2-way Anova with post-hoc Tukey test: significant main effect for inactivated area (F=10.78, p=0.0011).* ***D.*** *Same as C, but using a time window corresponding to the time a bar took to travel over a 10 deg window centered around peak response. P=0.0004, 2-way Anova with post-hoc Tukey test: significant main effect for inactivated area (F=11.10, p=0.0009).* ***E.*** *Same as Fig. 2.2E, but using the geometric compared to the arithmetic average. Error bars show the geometric standard error.*

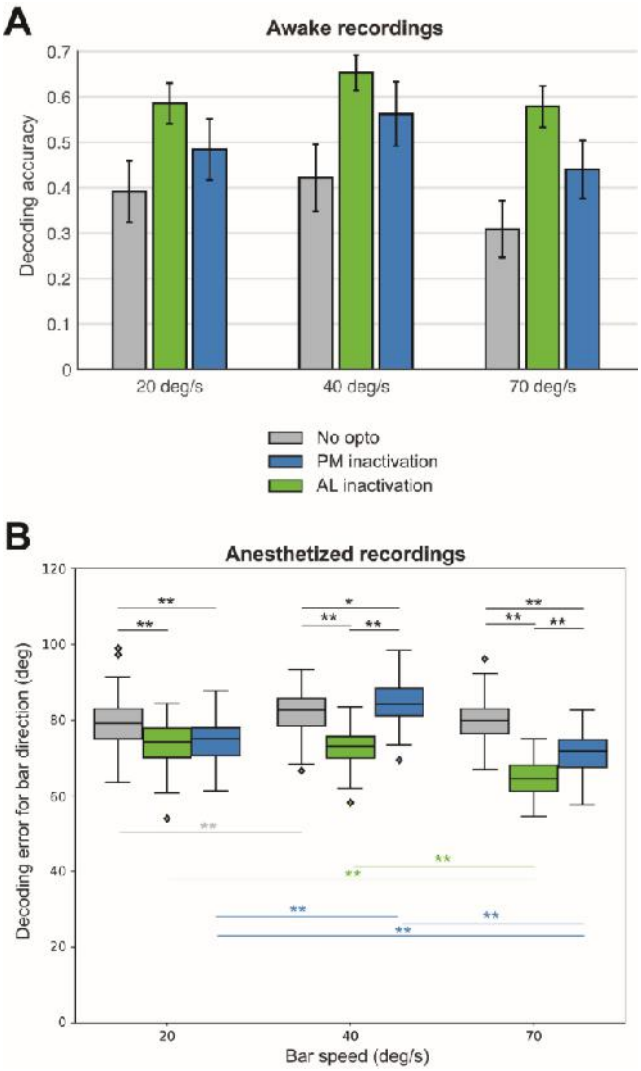


Supplementary Figure 2.8 – Optogenetic inactivation of area AL and PM under anesthesia asymmetrically reduces receptive field size of V1 neurons in the rising but not in the decaying phase of sensory evoked responses. ***A.*** *Same as Fig. 2.3A, but now for recordings performed under anesthesia.* ***B.*** *Same as Fig. 2.3B, but now for recordings performed under anesthesia (P-values for significant differences: 20 deg/s rising phase p=0.0003; 40 deg/s rising phase p=1.8x10^{-6}; 70 deg/s rising phase p=3.8x10^{-8}).* ***C.*** *Same as Fig. 2.3C, but now for recordings performed under anesthesia.* ***D.*** *Same as Fig. 2.3D, but now for recordings performed under anesthesia (P-values for significant differences: 20 deg/s rising phase p=0.0038; 40 deg/s rising phase p=6.9x10^{-6}; 70 deg/s rising phase p=0.0002). The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*

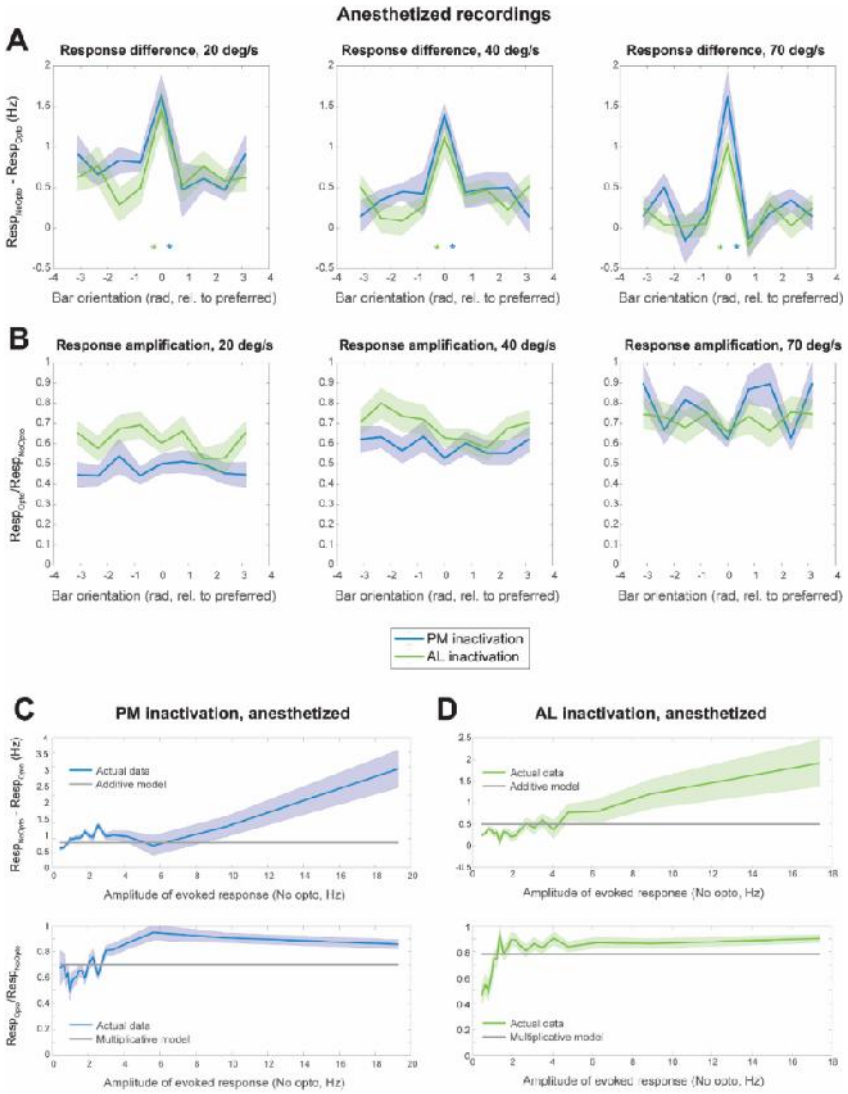


Supplementary Figure 2.9 – Optogenetic inactivation of area AL and PM under anesthesia enhances orientation and direction selectivity in V1. A. *Same as Fig. 2.4A, but now for recordings performed under anesthesia.* ***B.*** *Same as Fig. 2.4B, but now for recordings performed under anesthesia.* ***C.*** *Same as Fig. 2.4C, but now for recordings performed under anesthesia.* ***D.*** *Same as*

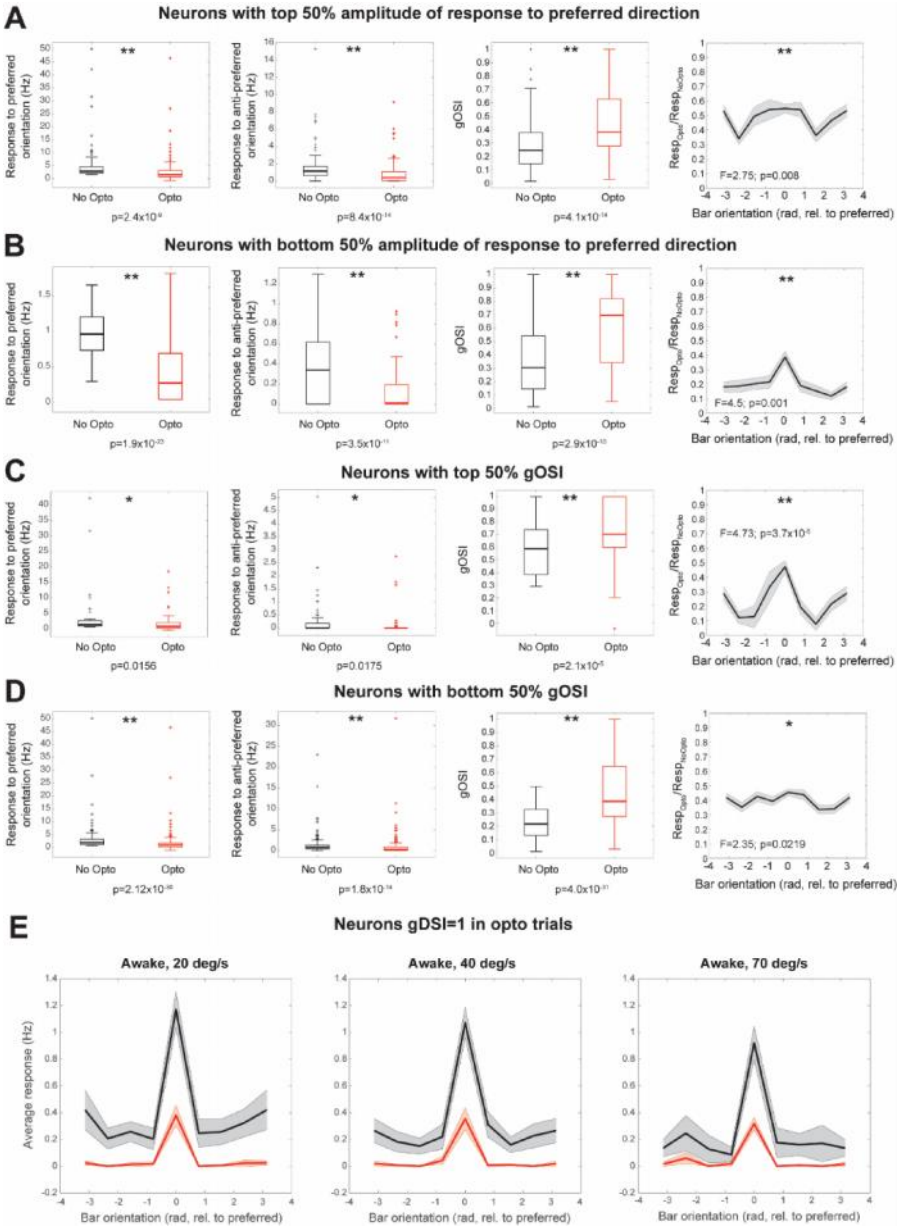
Fig. 2.4D, but now for recordings performed under anesthesia. ***E.*** *Same as Fig. 2.4E, but now for recordings performed under anesthesia. Asterisks indicate significant differences either between speeds (for a given area) or between inactivation of distinct areas, given the same speed ($p<0.05$, 2-way Anova with post-hoc Tukey test (for inactivation of area PM: $F=6.59$, $p=0.0016$)).* ***F.*** *Same as Fig. 2.4F, but now for recordings performed under anesthesia. Asterisks indicate significant difference either between speeds (for a given area), or between inactivation of distinct areas, given the same speed ($p<0.05$, 2-way Anova with post-hoc Tukey test (for inactivation of area PM: $F=19.01$, $p=2x10^{-8}$)). In neither panel E nor F were significant differences between areas (separately for each speed) found. The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*



Supplementary Figure 2.10 – Decoding accuracy during awake recordings; Optogenetic inactivation of area AL and PM under anesthesia enhances single-trial decoding of bar orientation in V1. A. *Decoding accuracy was computed as the proportion of bar directions correctly estimated on the same dataset used in Fig. 2.5. Accuracy is in line with previously shown values (Cai et al., 2018) and can be seen to increase upon inactivation of both PM and AL. Bars indicate mean ± SEM.* ***B.*** *Same as Fig. 2.5, but now for recordings performed under anesthesia. (Main effect for inactivacted area: F=227.9, p=0.0000; main effect for bar speed: F=178.0, p=0.0000; interaction effect: F=60.6, p=0.0000). The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*



Supplementary Figure 2.11 - Areas AL and PM enhance visual responses in V1 under anesthesia in accordance with neither an additive nor a multiplicative gain modulation. A. *Same as Fig. 2.6A, but now for recordings performed under anesthesia ($p<0.05$, one-way Anova with post-hoc Tukey test; for PM-20deg/s: $F=3.49$, $p=0.0011$; for PM-40deg/s: $F=4.22$, $p=0.0001$; for PM-70deg/s: $F=6.99$, $p=4.6x10^{-8}$; for AL-20deg/s: $F=3.00$, $p=0.0043$; for AL-40deg/s: $F=3.99$, $p=0.0003$; for AL-70deg/s: $F=6.05$, $p=8.8x10^{-7}$).* ***B.*** *Same as Fig. 2.6B, but now for recordings performed under anesthesia. No significant differences were observed.* ***C.*** *Same as Fig. 2.6C, but for recordings performed under anesthesia.* ***D.*** *Same as Fig. 2.6D, but for recordings performed under anesthesia. In panels C and D, low-amplitude responses are more strongly suppressed compared to high-amplitude sensory-evoked responses ($p<2x10^{-16}$, Chi-square goodness of fit test against a uniform distribution). The number of neurons included in the analyses shown in the figure is provided in Table 2.1.*



Supplementary Figure 2.12 – The effect of optogenetic inactivation of PM and AL on V1 responses remains the same when stratifying V1 neurons in different cohorts. A. *The effect of optogenetic inactivation on V1 neuronal responses to bars moving at 20 deg/s was separately assessed for neurons with the top 50% sensory-evoked responses to the preferred direction. Neurons were pooled together from all session (i.e., including the inactivation of both PM and AL). From left to right: responses to the preferred orientation in the absence (black) or presence (red) of optogenetic inactivation; responses to the anti-preferred orientation in the absence (black) or presence (red) of optogenetic inactivation; gOSI in the absence (black) or presence (red) of optogenetic inactivation;*

*amplitude modulation as a function of bar orientation. For the leftmost 3 plots, asterisks indicate significant differences between opto and no-opto conditions (paired t-test). For the rightmost plot, asterisks indicate if the amplitude modulation of sensory evoked responses was the same for all orientations (null hypothesis, corresponding to a flat line in the plot), or not (one way Anova). In all panels: *= p<0.05, **=p<0.001 ; exact p-values are reported in each panel. Similar results were found for responses to bars moving at 40 deg/s or 70 deg/s.* ***B.*** *Same as A, but for neurons with the bottom 50% sensory-evoked responses to the preferred direction.* ***C.*** *Same as A, but for neurons with the top 50% gOSI.* ***D.*** *Same as A, but for neurons with the bottom 50% gOSI.* ***E.*** *Same as Fig. 2.2B, but for neurons with gDSI=1 in the presence of optogenetic inactivation of either PM or AL. The number of neurons included in this analysis was 30 (20 deg/s), 34 (40 deg/s) and 39 (70 deg/s).*

Chapter 3: Multisensory task demands temporally extend the causal requirement for visual cortex in perception

Matthijs N. Oude Lohuis*
Jean L. Pie*
Pietro Marchesi
Jorrit S. Montijn
Christiaan P.J. de Kock
Cyriel M. A. Pennartz#
Umberto Olcese#

*These authors contributed equally
#These authors contributed equally

Published in:
Nature Communications (2022)
13:2864
10.1038/s41467-022-30600-4

Abstract

Primary sensory areas constitute crucial nodes during perceptual decision making. However, it remains unclear to what extent they mainly constitute a feedforward processing step, or rather are continuously involved in a recurrent network together with higher-order areas. We found that the temporal window in which primary visual cortex is required for the detection of identical visual stimuli was extended when task demands were increased via an additional sensory modality that had to be monitored. Late-onset optogenetic inactivation preserved bottom-up, early-onset responses which faithfully encoded stimulus features, and was effective in impairing detection only if it preceded a late, report-related phase of the cortical response. Increasing task demands were marked by longer reaction times and the effect of late optogenetic inactivation scaled with reaction time. Thus, independently of visual stimulus complexity, multisensory task demands determine the temporal requirement for ongoing sensory-related activity in V1, which overlaps with report-related activity.

Introduction

During perceptual decision making, stimulus presentation triggers an early response component in primary sensory cortices (driven by thalamic bottom-up input (Harris and Mrsic-Flogel, 2013)) and, often, a late component, thought to mostly result from recurrent activity through top-down, cross-areal interactions (Crochet et al., 2018; Cul et al., 2007; Supèr et al., 2001). Traditional accounts of how sensory stimuli are transformed into appropriate behavioral outputs have mostly characterized this process in terms of feedforward architectures, where progressively higher-order areas extract sensory features of increasing complexity (DiCarlo et al., 2012) to eventually instruct motor output. In the visual cortical system, a fast-acting (<150 ms) feedforward sweep is sufficient for image categorization (Thorpe et al., 1996). Accordingly, deep feedforward neural networks, inspired by this cortical hierarchical architecture, achieve near-human performance in image recognition (LeCun et al., 2015; Serre et al., 2007). The function of recurrent architectures has been primarily interpreted in the context of processing ambiguous or complex stimuli, for cognitive processes such as attention, and for consciousness (Dehaene and Changeux, 2011; Desimone and Duncan, 1995; Lamme and Roelfsema, 2000; Martínez et al., 1999; Mehta et al., 2000; Noesselt et al., 2002; Pennartz, 2015). For example, extra-classical receptive field effects in the visual system, such as surround suppression, and separating objects from background, are thought to depend on feedback projections from higher to lower visual areas (Allman et al., 1985; Boehler et al., 2009; Rao and Ballard, 1999; Roelfsema, 2006; Schnabel et al., 2018). Perceptual decisions involving figure-ground segregation require recurrent processing (Roelfsema, 2006), the duration of which becomes longer as a function of visual scene complexity (Kirchberger et al., 2021). Recently, a form of late activity in rodent V1 that reflects non-sensory variables such as movement, perceptual report, and arousal (Allen et al., 2017b; Kaplan and Zimmer, 2020; Musall et al., 2019; Salkoff et al., 2020; Steinmetz et al., 2019; Stringer et al., 2019) was shown to originate in prefrontal areas and progressively involve more posterior areas including sensory cortices (Allen et al., 2017b; Steinmetz et al., 2019).

Many hypotheses have been proposed on the function of late, recurrent activity in sensory cortices (including distributed motor command generation and context-dependent sensory processing)(Kaplan and Zimmer, 2020), but how it causally contributes to perception is debated. Across primates and rodents, the magnitude of late activity correlates with behavioral reports of perception (Cul et al., 2007; Manita et al., 2015; Sachidhanandam et al., 2013; Supèr et al., 2001). Suppressing late activity in the primary somatosensory cortex impairs tactile detection (Sachidhanandam et al., 2013), whereas in primary visual cortex it has been argued that feedforward activity is sufficient for visual discrimination (Resulaj et al., 2018; Thorpe et al., 1996). We hypothesize that the cognitive demands of a task, which are captured by the set of task rules as instantiated in an attentional set – see e.g. (Dias et al., 1996) – determine the temporal extension of the causal involvement of V1 in perceptual decision making, independently of stimulus complexity. The cognitive load may increase as a consequence of increasing attentional demands, such as when multiple sources of information need to be simultaneously monitored. For instance, integration of visual information with other sensory modalities (Meijer et al., 2019; Pennartz, 2015, 2009) may extend the time required by frontal and pre-motor regions to converge to a decision. This process might reflect an evidence accumulation model (Gold and Shadlen, 2007; Philiastides et al., 2006), i.e. a need to integrate information originating in V1 for longer periods in the case of complex, multisensory tasks. Analogously, the predictive

processing framework (Friston, 2005; Pennartz et al., 2019; Rao and Ballard, 1999) posits that visual and decision-related areas will keep on interacting via recurrent connections to jointly represent sensory stimuli and transform them into appropriate motor responses, performing computations for a time interval that may depend on task demands. Therefore, increasing the cognitive load required to perform a task, for instance by introducing the need to simultaneously monitor multiple sensory modalities, may extend the temporal window during which V1 activity remains causally required for perception, independently of visual stimulus features.

Results

To address this hypothesis, we trained mice in three versions of an audiovisual change detection task (task A) with the same stimulus configurations, but different reward contingencies. Head-fixed mice were presented with a continuous audiovisual stream of inputs with occasional instantaneous changes in the orientation of the drifting grating (visual trial) or the frequency of harmonic tones in a 'Shepard' stimulus (Shepard, 1964) (auditory trial, Fig. 3.1a-b; see Supplementary Fig. 3.1 for auditory stimulus details). We varied the amount of orientation change (visual saliency) and frequency change (auditory saliency) across each animal's perceptual threshold and fit all behavioral data according to a psychometric multi-alternative signal detection framework (Sridharan et al., 2014). We implemented three distinct task contingencies. First, for noncontingently exposed mice (NE, n=7) neither vision nor audition was predictive of reward, and these mice did not selectively respond to the stimuli (Fig. 3.1c). In a second version, only vision was associated with reward, and these unisensory-trained mice (UST, n=4) were thus trained to selectively respond to visual changes only, and ignore auditory changes (Fig. 3.1d). Third, multisensory-trained mice (MST, n=17) were trained to detect both visual and auditory changes (Fig. 3.1e; e.g. lick left for vision, lick right for audition). Phrased differently, all mice were presented with the same stimuli during training and testing, but lick responses to visual changes were only rewarded in UST and MST mice, and auditory changes only in MST mice. Trials with a stimulus change are indicated as "hit" when the mouse licked the correct spout and "miss" when no lick was provided; error trials correspond to trials with a lick towards the incorrect spout. To compare across cohorts, we also defined (surrogate) hit and miss trials for NE mice, based on whether (unrewarded) licks were performed after stimulus change. In all cohorts, mice performed many trials (mean 569, range 210-1047 per session). The discriminability index (d-prime) was high only for rewarded

contingencies, in both the auditory and visual modality (Fig. 3.1f; for individual mice, see Supplementary Fig. 3.2).

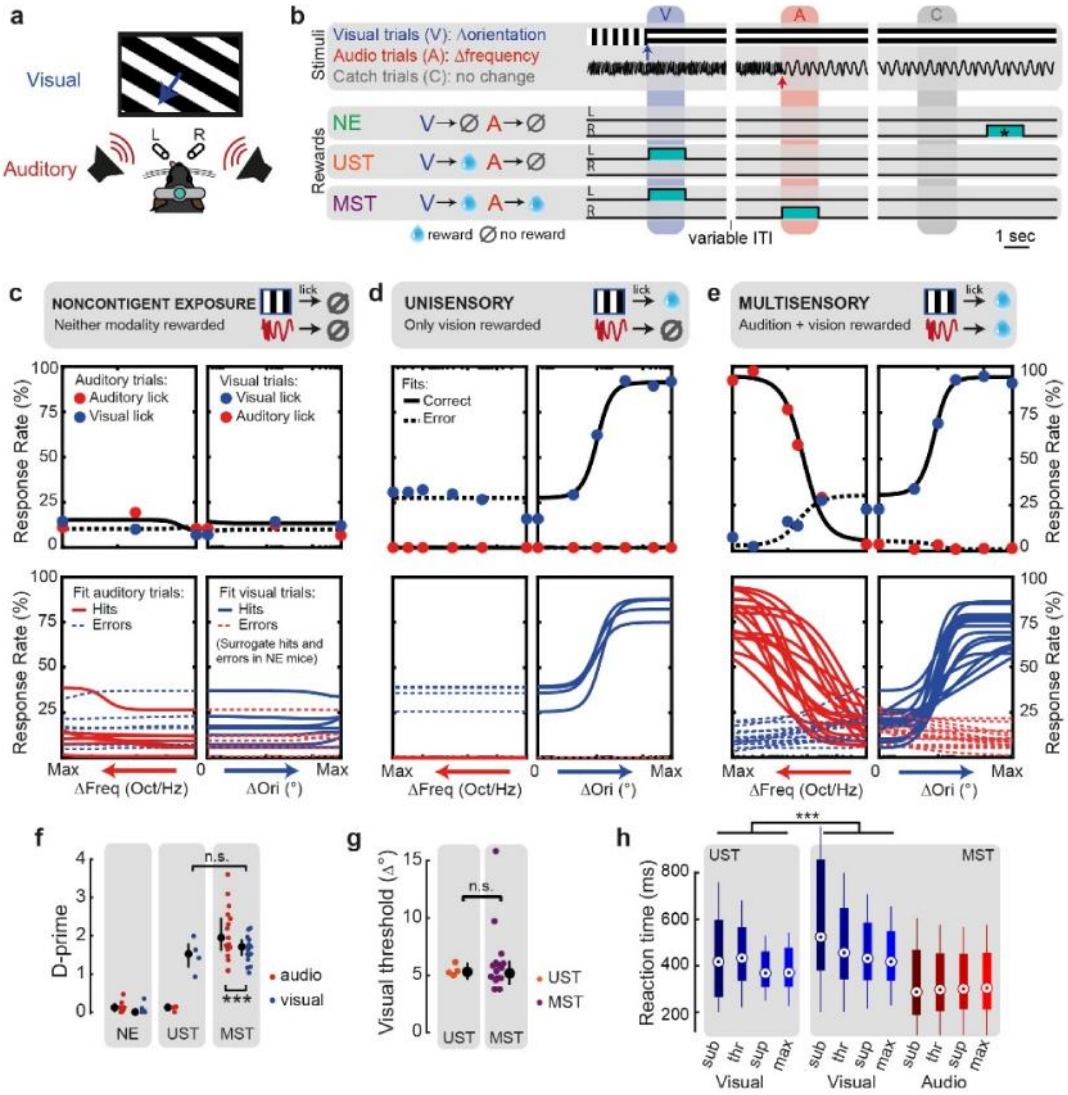


Fig. 3.1 Multisensory task contingencies delay reaction time. a *Schematic of task setup.* ***b*** *Example trial structure with reward availability for each cohort. Three cohorts of mice were presented with the same sensory stimuli: continuous drifting gratings that occasionally changed orientation and direction (visual trial) and a continuous tone that changed frequency content (auditory trial, Supplementary Fig. 3.1). Cohorts differed in reward structure. Noncontingently exposed (NE) mice were not rewarded contingently to the stimuli. Unisensory trained mice (UST) were rewarded for licks to the left spout after visual trials only, i.e. trained on vision only (cyan blocks denote the reward windows). Multisensory trained mice (MST) were rewarded and trained to lick (for instance) left to report visual changes and right to report auditory changes, i.e. discriminate modality. For NE mice, reward windows were temporally decorrelated from the sensory stimuli, and randomly occurring outside the stimulation period (these windows are denoted as cyan blocks*

with an asterisk). The trial windows indicate the time window used post-hoc to compare stimulus-related lick rates across cohorts; colors of these window correspond to the different trial types (blue: visual; red: auditory; grey: catch). For NE mice and auditory trials in UST mice, licks to the visual spout and auditory spout that happened to fall in these windows were defined as surrogate 'hits' and 'errors' (see Methods). ITI: inter-trial interval. ***c*** *The upper panels show behavioral response rates (dots) and model fits (lines: solid lines for responses to the correct – rewarded – side, dashed lines for responses to wrong – unrewarded - side) for an example session of a noncontingently exposed (NE) mouse. The bottom panels show the average psychometric fits for each mouse obtained by averaging parameters over sessions. Each session was fit with a two-alternative signal detection model (black lines in upper panels, colored in lower panels).* ***d*** *Same as (**c**), but for UST animals. Note how visual hit rates increase as a function of the amount of visual change, but not auditory change. The relatively high lick rate to the visual spout upon auditory changes arises because only that spout was associated with reward in this task.* ***e*** *Same as (**c**), but for MST animals. Hit rates increased as a function of both visual and auditory change.* ***f*** *D-prime across cohorts. Visual d-prime was comparable for UST and MST (ANOVA, n=151 sessions, $F(1,29) = 1.60$, $p = 0.22$,), and lower than auditory d-prime (ANOVA, n=139, $F(1,261) = 36.26$, $p=5.84x10^{-9}$). Each dot is the average over sessions for each animal. Error bars denote the median and interquartile ranges.* ***g*** *The detection threshold for visual orientation changes was comparable for UST and MST (ANOVA, n=151 sessions, $F(1,31) = 0.45$, $p = 0.51$).* ***h*** *Reaction time for the same subjectively salient visual stimuli (see Methods) was significantly shorter for UST compared to MST (ANOVA, n=3917 trials, $F(1,3865) = 60.1$, $p=1.11x10^{-14}$). Saliency levels: sub=subthreshold, thr=threshold, sup=suprathreshold, max=maximal change. Boxplot: dot, median; box limits, 25th and 75th quartiles; whiskers, 1 × interquartile range. $^{**}p<0.01$, $^{***}p<0.001$.*

Multisensory task contingencies delay reaction time

First, we wondered if visual performance was similar in the unisensory and multisensory task variants (UST and MST) and whether the more complex task contingency slowed responses. There were no significant differences between the cohorts for either maximum d-prime (Fig. 3.1f), discrimination threshold (Fig. 3.1g), or sensitivity (for statistics, see Supplementary Table 2.1). Reaction time, however, did vary across conditions (Fig. 3.1h). MST mice showed shorter auditory than visual reaction times and reaction times decreased with increasing levels of stimulus saliency for both UST and MST. For the same visual stimuli, reaction time was significantly longer for MST than for UST mice. For both vision and audition, reaction time negatively correlated with performance (Supplementary Fig. 3.2f, g). The addition of auditory task relevance thus increases reaction times for the same visual stimuli. This result was expected because MST mice were trained to make binary decisions on whether auditory versus visual changes took place, which requires comparisons across sensory channels (Hanson et al., 2009).

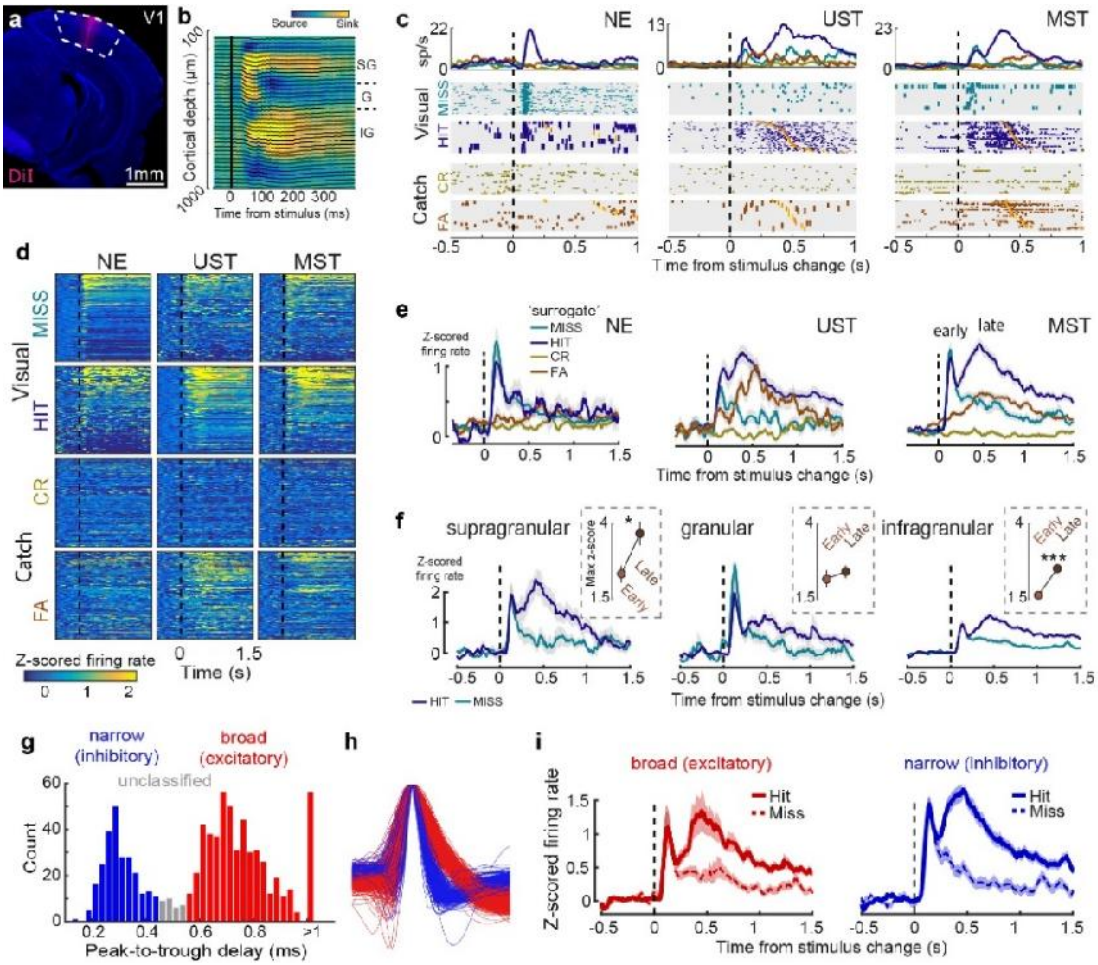


Fig. 3.2 Multisensory task demands modulate late activity in V1. ***a*** *Coronal histological section (3.6 mm posterior to bregma) showing silicon probe tract after recording in V1.* ***b*** *Representative example of current source density (CSD) map and LFP traces for checkerboard stimulation. SG=supragranular, G=granular, IG=infragranular. (CSD analysis was repeated with similar results for all 28 mice to determine the depth of probe insertion).* ***c*** *Raster plots (bottom) and firing rate (top panel) show stimulus-evoked and report-related activity in three example neurons recorded in V1 (Task A). Raster plots are grouped by trial type (visual or catch) and choice. Within trial type, trials are sorted by post-change orientation and response latency (orange ticks). Note that hits and misses in NE mice are surrogate conditions and are defined post-hoc. CR = correct rejection. FA = visual false alarm.* ***d*** *Heatmaps of trial-averaged z-scored activity of all neurons for the three cohorts for the same conditions as in (**b**). NE neurons: n=159; UST neurons: n=128; MST neurons: n=510.* ***e*** *Averaging z-scored firing rate over all neurons for visual and catch trials split by choice reveals biphasic activity in visual hits but not misses, with late activity only present in animals for which visual trials were rewarded (UST and MST). Note the increase in firing rates in FA trials for UST and MST mice but not NE mice. The weak early transient activity during FA trials in UST and MST mice during this noisy change detection task might be the result of stochastic variability interpreted as a sensory signal, i.e. falsely perceived changes(Kok et al., 2014), although a motor (lick) related signal cannot be excluded. See Supplementary Fig. 3.3 for an in-depth analysis of the lick-related nature of these responses.* ***f*** *Same as (**e**), but for each laminar zone*

(neurons from UST and MST mice combined). Inset: maximum z-score during the early (0-200 ms) and late (200-1000 ms) phase of visual hits (SG: F(1,194) = 4.60, p=0.03; G: F(1,171) = 0.00, p=1.00; IG: F(1,1284) = 23.32, p<0.001, ANOVA). ***g*** *Histogram of peak-to-trough delay for all neurons (n=816 neurons) colored by cell type class: narrow-spiking (peak-to-trough delay<0.45 ms; putative inhibitory; blue) and broad-spiking (peak-to-trough delay>0.55 ms; putative excitatory; red). Single units with intermediate peak-to-trough values were unclassified. The peak-to-trough delay was capped at 1 ms for neurons whose trough extended beyond the sampled window.* ***h*** *Normalized average waveform for all V1 neurons colored by cell type class.* ***i*** *Z-scored activity averaged over all broad-spiking V1 neurons (left, n=421 neurons) or narrow-spiking neurons (right, n=202 neurons) across UST and MST mice, split by hit/miss response for maximum visual change trials only. Throughout the figure, lines and shading are mean ± SEM.*

Early and late activity emerges in V1 of trained mice

To investigate whether delayed reaction times corresponded with slower dynamics of late V1 activity, we performed laminar recordings and sampled single-unit activity across cohorts (Fig. 3.2a). We used the current source density profile in response to visual stimulation and multi-unit activity profile to estimate recorded depth along the cortical layers (Fig. 3.2b). In NE animals the instantaneous orientation change evoked a short transient activity in V1 (until about 200 ms after stimulus onset) with a short-lasting tail (Fig. 3.2c). In visually trained animals (UST and MST), a similar transient wave occurred, but now also a late, second wave of activity was present (emerging around 200 ms after stimulus onset), primarily in hits and to a lesser extent in false alarms (Fig. 3.2d, e). These dynamics of early and late wave activity were seen for both threshold-level (thr) and maximal (max) orientation changes (Supplementary Fig. 3.3a, b). Splitting neurons based on recorded depth revealed different laminar dynamics. Early sensory-induced activity was most prominent in the granular and supragranular layers and was similar for hits and misses ($p>0.05$ for all laminar zones). During visual hits, late activity was prominent in supragranular and infragranular layers and was stronger than early activity (Fig. 3.2f). The late hit-related modulation (hits – misses, subtracted z-scored firing rate during 200-1000 ms) was stronger in supra- and infragranular layers than in the granular layer ($F(2,771) = 4$, $p=0.019$, ANOVA; Posthoc comparison: IG vs G: $F(1,784)=12.97$, $p<0.001$; G vs SG: $F(1,784)=6.50$, $p=0.01$; IG vs SG: $F(1,784)=0.58$, $p=0.45$. This is consistent with the idea that the granular layer is more strongly driven by thalamocortical afferents and extragranular layers more by recurrent processing. We also classified single units based on the delay between peak voltage deflection and subsequent trough. The histogram of peak-to-trough delay showed a bimodal distribution allowing clear classification into narrow and broad-spiking cell types (Fig. 3.2g, h). The dynamics of early and late components were present in both cell classes (Fig. 3.2i), suggesting a balanced increase in both excitatory and inhibitory activity upon hits.

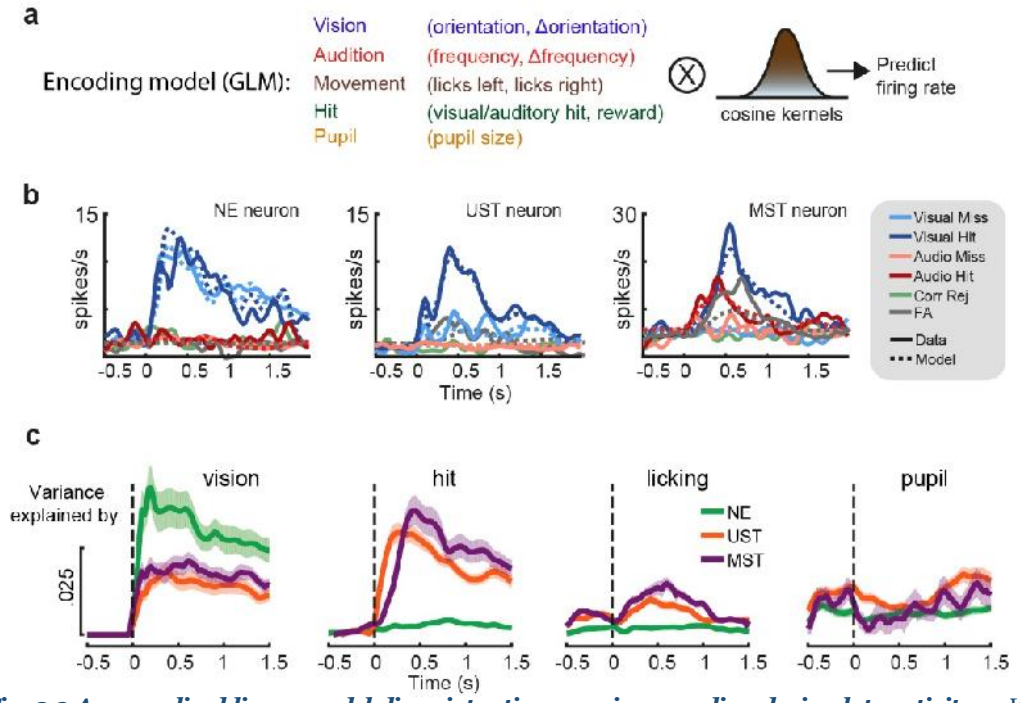


Fig. 3.3 A generalized linear model dissociates time-varying encoding during late activity. ***a*** *We constructed a kernel-based GLM encoding model in which variables related to the sensory environment, task, and behavioral state were included as predictors of firing rate. Binary task variables were convolved with raised cosine basis functions that spanned the relevant time window to model transient firing rate dynamics.* ***b*** *Model fits for three example neurons (one V1 neuron from each task version) show that predicted and actual firing rates closely overlap and that both sensory-driven activity (example #1), as well as report-related activity for both visual and auditory hits (examples #2-3), are captured by the model. The five stimulus-response combinations that had the most counts are plotted (trials with false alarms and licks to the incorrect spout are omitted).* ***c*** *Explained variance over time for subsets of predictors. Each line shows how much firing rate variance is explained for each time bin across trials based on only including a subset of all predictors. Shaded area corresponds to s.e.m.*

Neural coding during late V1 activity

Recent studies have shown that late activity in V1 can reflect movement-related variables (Musall et al., 2019; Salkoff et al., 2020; Stringer et al., 2019). We aligned population activity to the first lick after stimulus onset and found that spiking activity across many neurons was indeed modulated by licking movements, specifically in UST and MST mice (Supplementary Fig. 3.3c-j). The amplitude of this modulation was higher in trials with correct versus incorrect licks. To further disentangle the contribution of stimulus variables (visual and auditory features and amount of change), movement-related variables (the timing and number of lick responses), hits (visual and auditory hits), and arousal (pupil size), we built a kernel-based generalized linear model (GLM)(Park et al., 2014; Runyan et al., 2017; Steinmetz et al., 2019) where we included these variables as predictors of firing rate (Fig. 3.3a; see Supplementary Fig. 3.4a, b for model performance). The encoding model predicted firing rate dynamics of V1 neurons over time (Fig. 3.3b) and to investigate the contribution of each of the variables we computed the cross-validated variance of firing

rate explained over time by each of these subsets of predictors (Fig. 3.3c). In NE mice, visual predictors explained most of the variance with negligible contributions from other variables. In UST and MST mice, besides visual predictors, we found that both licking movement and hits explained a significant fraction of variance (Fig. 3.3c, see also Supplementary Fig. 3.4c-d). In sum, late V1 activity reflected a combination of visual-, movement- (licking), and hit-related variables, but only in trained mice.

Multisensory context delays the time course of late activity

To quantify in more detail how the delayed reaction time in MST mice was associated with temporal coding dynamics of single neurons we used a receiver operating characteristic (ROC) analysis (Green and Swets, 1966; Rossi-Pool et al., 2016). Across task versions, the ratio of neurons coding for visual features (grating orientation, and occurrence of a visual stimulus change – i.e. visual trials versus trials with no stimulus change, catch trials) was similar across cohorts (Fig. 3.4a). In UST and MST mice, however, visual report (i.e. visual hits vs. visual misses, henceforth hit/miss) was also encoded by a substantial fraction of neurons, in line with the averaged z-scored activity (Fig. 3.2e) and our regression model (Fig. 3.3). To understand at which time points visual features and hit/miss coding could be read out, we plotted the fraction of neurons that significantly coded for each of these variables over time (Fig. 3.4b, Supplementary Fig. 3.5). Temporal dynamics were strikingly similar across cohorts for sensory variables, while hit/miss coding appeared later in V1 for MST than UST mice. When we binned neurons based on their recorded cortical depth, we found that orientation coding was present across cortical layers and the coding of visual change occurrence was confined to an early transient wave in granular and supragranular layers (Fig. 3.4c). In contrast, hit/miss coding during late activity was predominant in infragranular layers. This spatial segregation with coding of visual change predominant in superficial layers and hit/miss coding in deeper layers suggests that these two processes have different neural substrates (see also Fig. 3.2f). We quantified the earliest moment of a significant increase in the fraction of coding neurons relative to baseline and found that only hit/miss coding was delayed in MST compared to UST (Fig. 3.4d; threshold changes: 288 ms ± 36 ms versus 162 ± 28 ms, MST vs. UST, $p<0.05$; maximal changes: 249 ± 104 ms vs 92 ± 56 ms, MST vs. UST, n.s.). In relation to the delayed visual reaction times in MST mice, we found that the onset of hit/miss coding correlated with reaction time at the level of population-averaged firing rate differences per session (Fig. 3.4e), as well as the bootstrapped estimate from single neurons across sessions (Fig. 3.4f). This result was also confirmed by a GLM-based analysis (Supplementary Fig. 3.4e). Hit/miss-related activity preceded the first lick by about 280 ms (Fig. 3.4f, Methods). Therefore, at 200 ms after

stimulus onset (blue dotted line in Fig. 3.4f) UST mice generally showed hit/miss coding in V1, while MST mice did not.

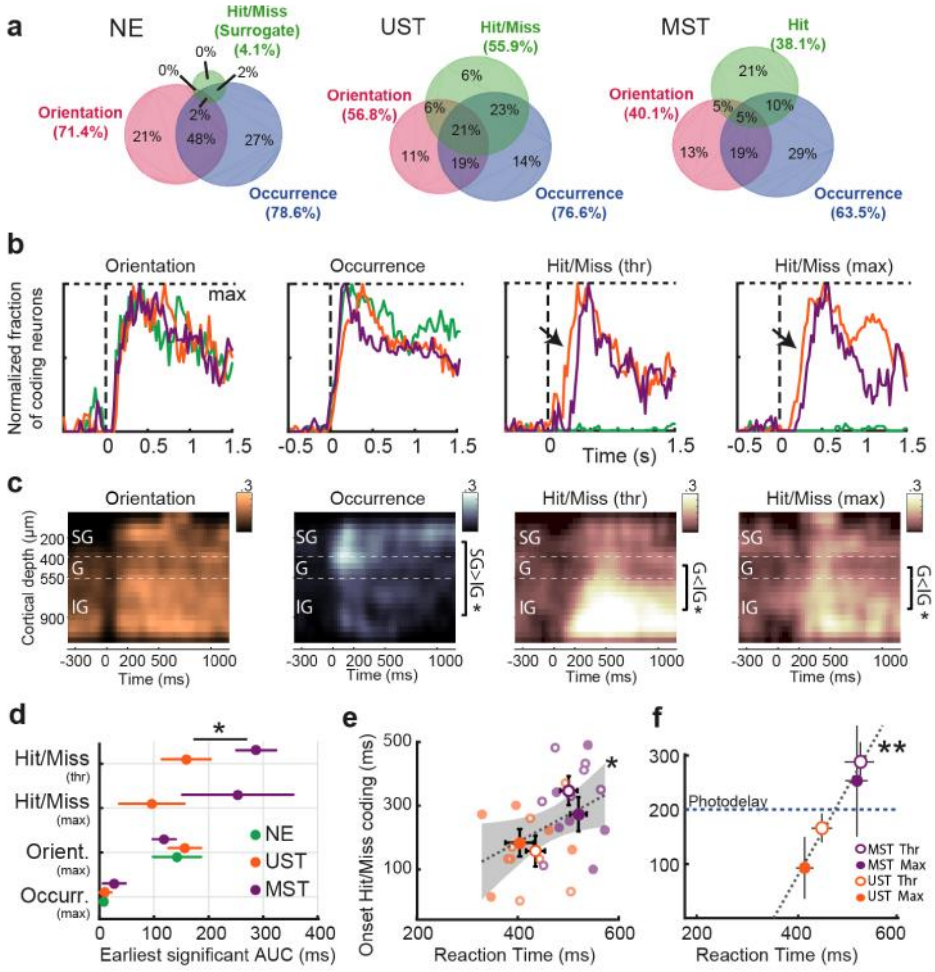


Fig. 3.4 The onset of late activity is delayed in MST mice. ***a*** *The Venn diagrams show for each training cohort the percentage of neurons encoding orientation (grating after stimulus change), occurrence (presence of a visual change or not), or hit/miss (visual hits versus visual misses, with no lick response) as established with ROC analysis. Only maximum visual change trials were used. Shown are percentages out of all coding neurons; percentage of non-coding neurons per cohort: NE: 15.5%; UST: 13.3%, MST: 35.6%.* ***b*** *Fraction of neurons (summed over all recordings) coding for task-relevant variables over time. Each coding fraction is baseline-subtracted and normalized by its maximum. Visual hit/miss coding (hits vs misses) was only present in UST and MST mice (as expected) and started earlier in UST than MST mice (highlighted with black arrows).* ***c*** *Heatmaps of the fraction of coding neurons across time and cortical depth, with neurons binned based on their recorded depth relative to the granular layer (400-550 µm from dura). Only UST and MST cohorts were included to compare sensory and hit/miss coding in the same datasets. SG=supragranular, G=granular, IG=infragranular. Occurrence coding, 0-200 ms, ANOVA, SG versus IG, $F(1,16)=7.21$,*

*p=0.016. Hit/miss coding, Thr, 200-1000 ms, G vs IG, F(1,15)=5.21, p=0.037; Max, G vs IG, F(1,15)=4.96, p=0.042. Significance (sidebars): * p < 0.05.* ***d*** *Earliest increase in the fraction of significantly coding neurons. Mean ± 95% CI (bootstrap). The apparent fast onset of visual occurrence coding is likely due to temporal smoothing of firing rates. (Bootstrap test, two-sided, UST n=128, MST n=306 neurons, p=0.012)* ***e*** *Reaction time correlated with the onset of hit/miss coding in population-averaged activity (ANOVA, n=26 sessions, F(1,24)=5.15, p=0.03). Grey dotted line shows linear regression fit. Each dot is one session. Error bars show mean ± SEM.* ***f*** *Same as (**e**), but now for the bootstrapped average for each visually trained condition using single neuron AUC (as in d). Reaction time correlated with the earliest moment of significant hit/miss coding (ANOVA, F(1,2)=102.33, p=0.0096). Error bars show bootstrapped mean and 95% CI. Blue dotted line at 200 ms marks the time point where late photostimulation was applied (see Fig. 3.5d). At this point, unisensory trained mice already showed hit/miss- coding in V1, while multisensory trained mice did not.*

Late activity is causally required for perceptual decision making

Next, we wondered whether V1 activity occurring after the onset of report-related activity could be causally linked to perception. We locally expressed ChR2 in parvalbumin-expressing interneurons in V1 (Fig. 3.5a, b, Supplementary Fig. 3.6a) to achieve temporally specific optogenetic inactivation (Madisen et al., 2010; Olcese et al., 2013). Laser stimulation robustly silenced multiunit activity (Fig. 3.5c). To determine the temporal window of V1 involvement, we silenced it either from the moment of stimulus change onwards ("Early", 0 ms) or from the 200 ms temporal cutoff we identified in the onset of hit/miss coding in UST and MST mice ("Late", 200 ms; Fig. 3.4f). Photostimulation was performed during a subset of all trial types (including catch trials, probing the effect of photostimulation without any relationship to stimuli or motor responses) and continued until the animal made a choice. Early blue light stimulation during visual trials (i.e. starting at the onset of stimulus change) reduced the activity of putative excitatory neurons to about 5% of their baseline activity. Late photostimulation left the initial sensory-driven response intact but silenced activity after 200 ms relative to stimulus onset (Fig. 3.5d).

Early silencing of V1 strongly reduced detection of orientation changes during both UST and MST task performance (Fig. 3.5e, g), consistent with the primary role of V1 in visual feature processing (Glickfeld et al., 2013b; Resulaj et al., 2018; Zatka-Haas et al., 2021). For threshold levels of orientation change (Fig. 3.5g), the detection of visual change was fully suppressed, indicating that early inactivation of V1 is sufficient to impair visual change detection. Interestingly, late silencing only affected change detection performance of MST mice (Fig. 3.5f, h). V1 silencing did not affect auditory change detection (Supplementary Fig. 3.6b). Moreover, photoillumination of control area S1 did not affect visual or auditory performance (Supplementary Fig. 3.7).

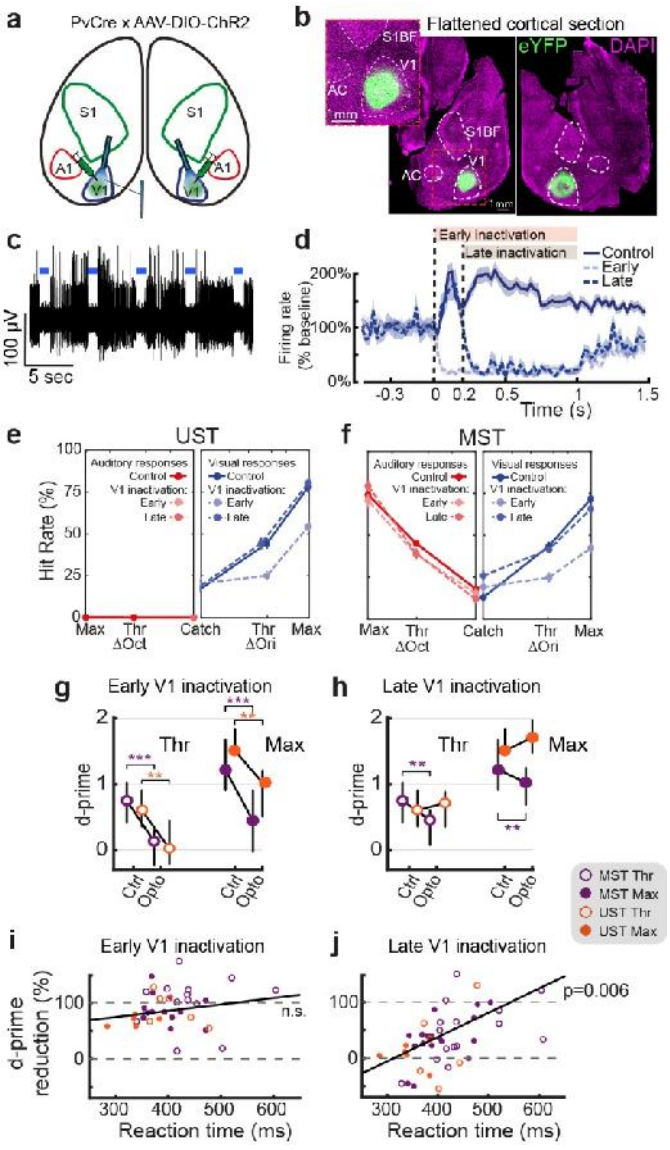


Fig. 3.5 Late silencing of V1 selectively impairs task performance of sessions with slow reaction time. a *Cre-dependent ChR2 expression in bilateral V1 of PvCre mice allowed robust silencing by locally enhancing PV-mediated inhibition. A1 = auditory cortex, S1 = primary somatosensory cortex.* ***b*** *Dorsal view of flattened cortical hemispheres sectioned approximately through layer 4 showing localized viral expression in bilateral V1. (Repeated with similar results for all 28 mice)* ***c*** *High-pass filtered trace from an example V1 recording site showing robust silencing of multi-unit spiking activity during bouts of 1-second photostimulation (blue bars).* ***d*** *Baseline-normalized firing rate averaged over V1 neurons from UST and MST mice. Control trials are visual hits. Mean ± SEM.* ***e*** *Behavioral response rates for control, early, and late silencing trials follow plotting conventions of Fig. 1c-e.* ***f*** *Same as (**e**), but for MST mice. Both early and late silencing affected visual change detection rates. For the increase in FA see Methods.* ***g*** *Early silencing affected visual discrimination performance (d-prime) for both saliencies across UST and MST cohorts. (ANOVA, UST n=18, MST n=34 sessions, UST Thr, $F(1,32)=16.71$, $p=0.0032$; UST Max, $F(1,32)=14.80$, $p=0.0064$; MST Thr, $F(1,59)=35.32$, $p=2\times10^{-6}$; MST Max, $F(1,58)=32.56$, $p=5\times10^{-6}$, each corrected for 4 multiple comparisons (Bonferroni-Holm))* ***h*** *Effect of late silencing depended on task type: late silencing only reduced d-prime in MST (same n as **g**, ANOVA, Thr, $F(1,54)=13.90$, $p=0.00553$, Max, $F(1,53)=13.48$, $p=0.0067$), but not UST mice (Thr, $F(1,32)=0.29$, $p=1$, Max, $F(1,30)=1.19$, $p=0.85$). For both **g** and **h**, *p<0.05, **p<0.01, ***p<0.001, errorbars denote inter-quartile range.* ***i*** *The effect of early silencing (quantified as the reduction in d-prime) was not significantly correlated with the median reaction time in control trials from the same session (ANOVA, n=40 conditions, $F(1,33)=1.71$, $r=0.048$, $p=0.865$).* ***j*** *Same as (**i**) but for late silencing. The effect of late silencing was significantly correlated with the reaction time ($n=45$, $F(1,15)=10.04$, $r=0.423$ $p=0.03$).*

Even though late silencing impaired visual change detection in MST mice *on average*, results across animals and experimental sessions were mixed: some sessions showed robust behavioral impairment, whereas others showed little effect (Supplementary Fig.

3.6c, f). We hypothesized that this variability could relate to changes in the speed of the perceptual decision-making process, with sessions having slow reaction times being proportionally more affected due to extended reliance on V1. In fact, this could be observed when we separately analyzed the behavioral consequences of optogenetic inactivation on the fastest and slowest sessions (top and bottom 50% of all sessions split by mean reaction time – see Supplementary Fig. 3.6c-h). To more precisely quantify this, we plotted the reduction in d-prime as a function of reaction time, the latter quantified in control trials within that same session (as a proxy for behavioral reactivity in the same animals) (Fig. 3.5i, j, Supplementary Fig. 3.6i, j). Whereas early silencing invariably reduced performance, the effect of late silencing scaled with reaction time (Fig. 3.5j), and amounted to a complete impairment in visual detection (100% reduction) for sessions with the longest RTs. This impairment was not found for the animal's propensity to lick (which affects false alarms and was quantified with the criterion parameter in our behavioral model; Supplementary Fig. 3.6k), suggesting that the effect of late inactivation on perceptual sensitivity increased as a function of rising RT. Furthermore, we considered the possibility that the reduction in d-prime relative to control trials was confounded by a lower d-prime on control trials for slow sessions to begin with (Supplementary Fig. 3.2g, see also Supplementary Fig. 3.6l), but this did not account for the effect (Supplementary Table 3.1). Late silencing thus left performance intact in 'fast' sessions in which hit/miss coding emerged quickly (mostly UST, but also some MST sessions) and reduced performance in slow sessions where hit/miss coding started after 200 ms (mostly MST sessions with higher cognitive demands, but note also how one slow UST session was affected; Fig. 3.5j).

Causal involvement of late activity generalizes to visuotactile side detection

So far, our results suggest that in the multisensory variant of the change detection task (MST), late V1 activity is causally involved whereas in the unisensory variant it mostly is not. However, UST and MST cohorts do not only differ by sensory contingencies, as UST mice were trained on a Go/No-Go paradigm, while MST mice learned a two-alternative choice task. Thus, the results we report could be due to differences in behavioral strategy rather than to changes in multisensory context. Furthermore, we wondered whether our results may extend to other sensory modalities. To address these aspects, we developed a visuotactile side detection task in which mice reported the side of sensory stimulation, i.e. instructing them to lick left for visual or tactile stimuli presented to the left and oppositely for the right side (Task B; Fig. 3.6a). Stimuli consisted of monocular drifting gratings (visual), whisker pad deflection (tactile), or a combination of both. In this task B, visual and tactile information need to be integrated as an inclusive-OR operation (rather than discriminated as in task A) to decide on which side (left/right) sensory stimuli appeared. Again, some mice were trained on responding only to vision to obtain reward (UST), while another cohort was trained on both vision and somatosensation (MST). Importantly, this UST version contained two response options and required responding to the correct lick spout (the visual stimulus could appear on the left or right). In addition to the differences with task A, this new task allowed us to test if our results extended to another multisensory processing principle (congruent combination of modalities instead of segregation(Meijer et al., 2019)) and the detection of a different stimulus dimension (contrast instead of orientation change).

We controlled stimulus salience by varying visual contrast and whisker deflection amplitude and fitted the behavioral data with a psychometric model (Fig. 3.6b; see

Methods). The visual detection threshold and task performance at maximum saliency were similar for both cohorts (UST and MST, Supplementary Fig. 3.8a, b). As in Task A, we found that visual reaction time (RT) decreased for higher stimulus saliencies (Supplementary Fig. 3.8c). In contrast with task A, however, RTs were similar between tactile and visual trials.

Pursuing the comparison with task A, laminar recordings in V1 revealed similar neural dynamics, with a marked early stimulus-driven component visible in contralateral visual trials, and late activity in both contra- and ipsilateral visual hits (Fig. 3.6c, Supplementary Fig. 3.8f; see also Supplementary Fig. 3.8g for lick-aligned neuronal activity). This late V1 activity was also present in tactile-only hits – so without visual stimuli - although only for MST mice (Supplementary Fig. 3.8h). We optogenetically silenced unilateral V1 either from stimulus onset (early silencing) or after a delay that separated the late from the early wave of activity (240 ms, late silencing; Fig. 3.6c). Early silencing of unilateral V1 reduced the detection of contralateral threshold-contrast stimuli in both UST and MST mice but spared detection of ipsilateral stimuli and full contrast stimuli (Fig. 3.6d). In MST mice, tactile detection was not affected by V1 silencing (Supplementary Fig. 3.8d). Consistently with our results for Task A, while early V1 silencing impaired detection of threshold-level visual stimuli for both unisensory and multisensory contexts, late V1 silencing only affected such detection in MST mice (Fig. 3.6d, Supplementary Fig. 3.8e). As in Task A, we observed that the effect of silencing increased for conditions with longer reaction time (Fig. 3.6e). Overall, results obtained with task B generalize our findings and confirm that the temporal window for the causal involvement of V1 in perceptual decision making is extended when subjects reinstate the more demanding, multisensory attentional set they have been trained on.

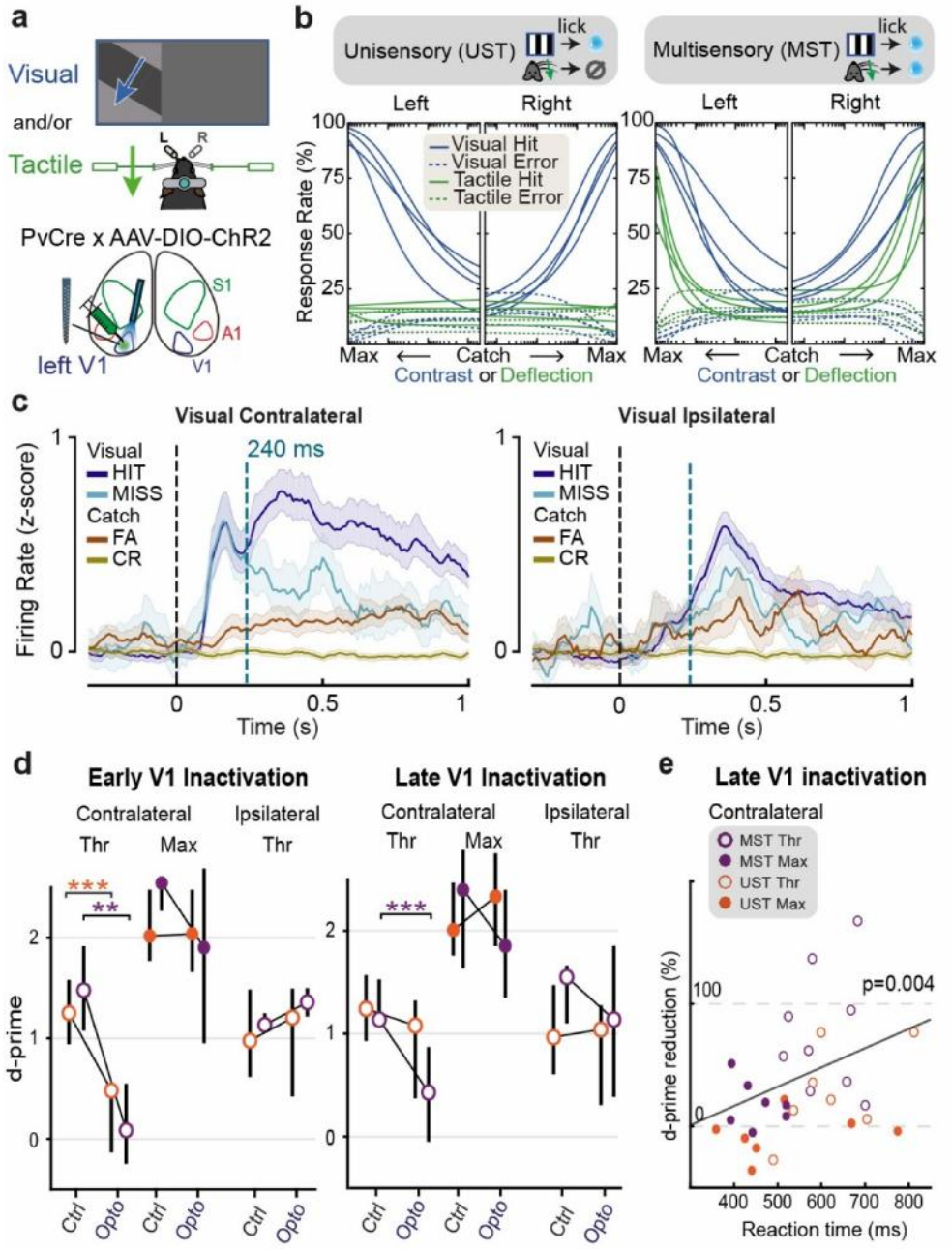


Fig. 3.6 Extended causal requirement of V1 generalizes to visuotactile side detection. ***a*** *Schematic of the visuotactile two-sided detection task in which mice reported the side of visual and/or tactile stimulation.* ***b*** *Psychometric fits for visual and tactile detection for each mouse trained in the UST and MST version of the task. Same conventions as in Fig. 3.1 c-e, with the x-axis running left for visual/tactile stimuli presented to the left side and right for visual/tactile stimuli presented to the right side. Visual and tactile intensities were normalized for rendering purposes.* ***c*** *Average z-scored firing rate of responsive V1 neurons during visual (max contrast) and catch trials, split by trial*

outcome. Only neurons from MST mice are shown. The dashed blue line indicates the onset of optogenetic silencing. Similarly to Task A, contralateral hits elicit more late activity compared to misses. Also note the weak early transient activity during FA trials (analogously to Task A, see Fig. 3.2e). Shaded area: bootstrapped 95% confidence intervals. ***d*** *Visual d-prime across V1 inactivation conditions. Early V1 inactivation (left) impaired the detection performance of contralateral threshold-level visual stimuli in both UST (ANOVA, n=7 sessions, $F(1,14)=24.57$, $p=0.0006318$) and MST (ANOVA, n=6, $F(1,12)=17.93$, $p=0.0023$), without effect on ipsilateral threshold-level stimuli nor contralateral maximum-level stimuli. Late silencing (right) only affected detection performance of contralateral threshold stimuli in MST mice (ANOVA, n=7 sessions, $F(1,14)=45.14$, $p=0.000036$), but not in UST mice (ANOVA, n=7, $F(1,14)=2.15$, $p=0.164$). $^{**}p<0.01$, $^{***}p<0.001$. Each corrected for 4 multiple comparisons (Bonferroni-Holm). Errorbars denote inter-quartile range.* ***e*** *Reduction in d-prime by late silencing correlated with the median reaction time on corresponding control trials (ANOVA, n=30 conditions, $F(1,26)=9.785$, $p=0.00427$, $r^2=0.7056$). Each dot represents a session.*

Population decorrelation during hit trials precedes and is locked to reaction time

The late report-related wave of activity during visual hits (Fig. 3.2e, Fig. 3.6c) likely arises through an interplay of higher-order areas that feed back to V1, possibly including premotor or other frontal areas (Allen et al., 2017b; Musall et al., 2019; Salkoff et al., 2020; Zagha et al., 2013). The timing of this wave predicts the behavioral effects of late V1 inactivation, but the underlying mechanism governing the sculpting of a behavioral decision remains unclear. One possibility is that late activity is predominantly related to movement variables, coded orthogonally to sensory representations from a population perspective (Stringer et al., 2019). To investigate this, we further explored the properties of late activity. First, we tested whether the extended causal requirement of V1 was related to changes in the fidelity of sensory processing, as indexed by orientation decoding. For the audiovisual change detection task, we observed that, in line with the overall neuronal population (Fig. 3.2), orientation-selective neurons also showed a late hit-modulation of firing rate (Fig. 3.7a) and examined the effect of report-related activity modulation at a population level by training a random forest decoder to decode post-change grating orientation from V1 population activity. Orientation decoding was possible for hundreds of milliseconds after the orientation change with comparable performance across the three task versions (Fig. 3.7b). This suggests that in all visual trial types (regardless of task contingencies) information regarding the orientation of the stimulus was similarly present and that the extended requirement of V1 could rather be due to the interaction of this representation with the rest of the cortical circuit. Correlated firing rate fluctuations that are unrelated to signal coding (noise correlations – NCs) can impact information coding in populations of neurons (Kohn et al., 2016; Montijn et al., 2016; Zohary et al., 1994). NCs decrease as a function of various conditions, for instance when animals become experts at change detection (Ni et al., 2018) or through attention (Cohen and Maunsell, 2009). We computed pairwise firing rate fluctuations over time for visual hits and misses. During

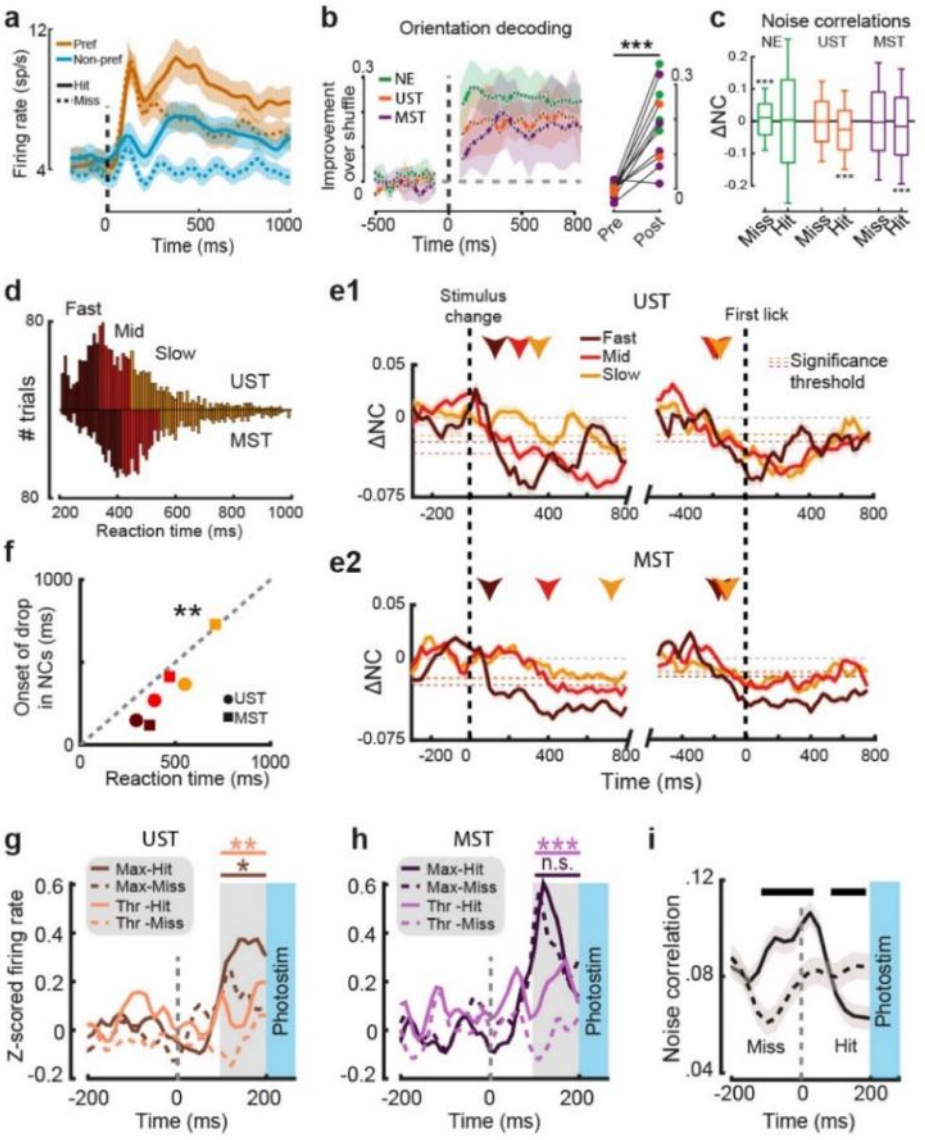


Fig. 3.7 Onset of report-related activity in task A and drop in noise correlations predict effects of late silencing. a *Average spiking rate for all orientation-selective neurons for preferred and non-preferred orientations split by hits and misses (UST and MST neurons combined; task A).* ***b*** *Orientation decoding performance over time. Right panel: decoding performance increased post-stim (0 to +500 ms) versus pre-stim (-500 to 0 ms; n=11 sessions, cohorts combined, ANOVA, $F(1,17)=44.76$, $p= 4.1x10^{-6}$) and increased in individual sessions from all cohorts (colored dots).* ***c*** *Change in noise correlation (NC) relative to baseline (200 to 1000ms compared to baseline -500 to 0 ms) for visual trials split by choice and cohort (for auditory trials, see Supplementary Fig. 3.9a). Boxplots show the median and interquartile range (box limits) and 0.5 x interquartile range (whiskers). Noise correlations decreased only during hits in UST and MST mice (ANOVA, UST, n= 1930 pairs, $F(1,3856)=82.44$, $p<1x10^{-19}$; MST n=13972, $F(1,28188)=142.96$, $p<1x10^{-33}$). Misses in NE mice were associated with a slight increase in noise correlations (n=2904 pairs,*

*F(1,5805)=14.67, p<0.001). **d** Reaction time distributions for visual hits in UST and MST cohorts and tertile ranges. **e1** Noise correlations over time with respect to baseline, either aligned to stimulus change (left) or first lick (right). Horizontal dashed lines indicate for each tertile the threshold for each tertile for the onset of the drop in NCs (below 2 standard deviations of the baseline; -500 to 0 ms) and this onset is highlighted with colored arrows. Note how noise correlations (aligned to stimulus change) drop first in fast trials, and progressively later in medium and slow trials. Right panels show that, when aligning to lick onset, the drop in noise correlations precedes reaction times by a similar lag, independent of reaction time tertile. **e2** Same as (**e1**), but for MST mice. **f** Reaction time and moment of decorrelation were significantly correlated (Pearson correlation, n=6, r=0.960, p=0.002). Scatterplot shows median reaction time and earliest time point of decorrelation for each tertile in the two visually trained cohorts. **g** Average Z-scored firing rates just before photostimulation (100-200 ms) were higher if the trial resulted in a visual hit rather than a miss in UST mice (Thr: F(1,156)=10.16, p=0.002; Max: F(1,152)=5.66, p=0.019; ANOVA). **h** Same as (**g**), but for MST mice. Firing rates just before photostimulation were higher for hits than misses only for threshold visual changes (F(1,268)=13.19, p=0.001), but not maximal changes (F(1,254)=0.59, p=0.44). **i** Noise correlations for visual hit and miss trials before photostimulation onset (grouped across UST and MST cohorts and saliency levels). Black bar on top indicates time bins with a significantly different NCs between hits and misses (p < 0.05). Throughout the figure, lines and shading are mean ± SEM.*

baseline (-500 to 0 ms) NC values were comparable to the literature (Cohen and Kohn, 2011; Ecker et al., 2010; Hansen et al., 2012; Meijer et al., 2020) (0.063 ± 0.14 std) and decreased after stimulus change only during hits in UST and MST but not NE mice (Fig. 3.7c, Supplementary Fig. 3.9a). To investigate whether the onset of the decorrelation was related to reaction time, we split all visual hits from V1 recording sessions into three tertiles based on reaction time (Fig. 3.7d). Similar to behavioral data without recordings (Fig. 3.1h), UST mice reacted faster than MST mice (p=0.0041, Wilcoxon rank sum test). We quantified the earliest time-point where the drop in NCs reached significance (relative to baseline) for each tertile for UST and MST mice. NCs decreased at a latency that depended on reaction time, with the drop in NCs occurring later on slow compared to fast trials (Fig. 3.7e). The latency of the decrease in NCs and reaction time were significantly correlated (Fig. 3.7f), suggesting that population decorrelation is time-locked to reaction time. Indeed, noise correlations relative to the first lick (see Methods) dropped just preceding this first lick, irrespective of reaction time (Fig. 3.7e right part, Supplementary Fig. 3.9b), with the strongest decrease for visual hits in UST and MST mice and no decrease in surrogate hits in NE mice (Supplementary Fig. 3.9c). Overall, similarly to the onset of late, report-related activity, a late drop in NCs also precedes and is time locked to perceptual report.

Activity level and decorrelation predict the effect of late silencing

If the late-onset increase in spiking activity and drop in NCs were related to perceptual report, one would expect that the variability in the behavioral effect of silencing (i.e. whether V1 inactivation was followed by a hit or miss) could be explained at the single-trial level by whether this drop had already occurred at the time of photostimulation. We therefore focused on V1 activity during visual trials just before late inactivation started (n=230 cells), and tested whether the report-related firing-rate modulation and drop in NCs had both already occurred before 200 ms in hits but not misses. Indeed, hits were associated with increased activity just before photostimulation started (100-200 ms after stimulus onset) across levels of stimulus change and task versions (Fig. 3.7g, h). Similarly, NCs showed distinct profiles for hits and misses (Fig. 3.7i) and decreased just before silencing onset during hits but not misses. Surprisingly, NCs were higher just before and after stimulus change on hit trials (-125 to +25 ms around stimulus change), in contrast

with (Beaman et al., 2017). Overall, these results show that increased firing rates, and the temporally coinciding drop of NCs in V1, correlate with and predict sensory report (Fig. 3.8).

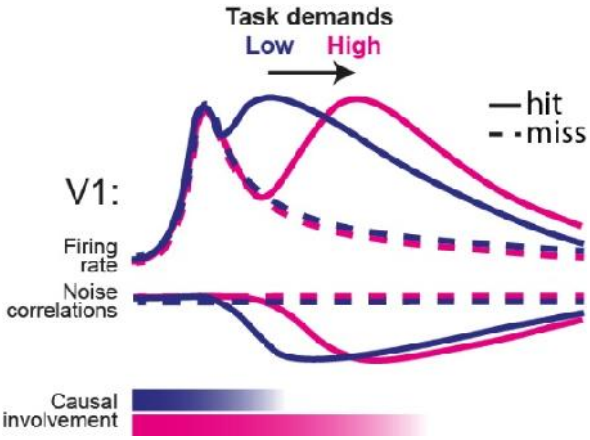


Fig. 3.8 Schematic summary of results. *Increased task demands (in our tasks imposed by multisensory requirements) delay the onset of the late report-related wave of activity and drop in noise correlations, and extend the causal involvement of V1. Jointly, these processes predict the behavioral effect of late V1 inactivation on visual detection, and whether a trial is going to be a hit or a miss.*

Discussion

In this study, we investigated the nature and function of late, recurrent activity in V1 in perceptual decision-making. An increase in multisensory task demands delayed behavioral decisions and extended the temporal window in which V1 was causally involved in determining perceptual report, given the same visual stimulus. As animals in the MST tasks were trained to process behaviorally relevant signals in two sensory modalities rather than one, longer reaction times (compared to UST tasks) are likely needed to integrate and compare information from distinct sensory modalities, also to assess which of two modalities is most likely to present an externally (as opposed to self-) induced sensory change. In particular, sensory selection was previously shown to affect neuronal processing already at the level of primary sensory cortices (Lee et al., 2016), via a thalamocortical feedback loop (Wimmer et al., 2015). Similarly, in our tasks the detection of a behaviorally relevant signal might modify how a given sensory modality is routed and processed (Meijer et al., 2019), leading to an extended requirement for V1 in multisensory contexts. Furthermore, in our experiment, a specific visual stimulus under conditions of low saliency (i.e., a small change in grating orientation) required more time to determine whether there was a change than in the case of high saliency (Fig. 3.1h, 3.4f). Therefore, our results may not be specific to multisensory contexts (Fig. 3.5j). Indeed, increased cognitive demands (implemented in terms of memory-dependent vs. visually-guided navigation) were shown to broaden the contribution of dorsal cortical areas to perceptual decision-making beyond V1 (Pinto et al., 2019). In conditions of high task demands, these and the present results suggest that higher-order cortices may need to rely on the output of V1 for longer temporal windows compared to simpler (e.g., purely visually guided) tasks. Importantly, we obtained similar results in two completely independent experimental groups (Task A and B). This further supports the existence of a link between task demands and temporal extension of V1 requirement.

We found modest differences in onset latencies and orientation coding of visually evoked V1 responses across task contingencies and cohorts, suggesting that the dynamics of bottom-up, feedforward processing are mostly conserved. In contrast, striking differences

were found in the late phase of V1 activity, and particularly in the behavioral consequences of late optogenetic inactivation. Previous studies in humans and non-human primates showed that attention-related activity in V1, similarly to report-related activity, emerges in a late phase of V1 sensory-evoked responses as a consequence of top-down modulation (Boehler et al., 2009; Martínez et al., 1999; Mehta et al., 2000; Noesselt et al., 2002; Supèr et al., 2001). A potential role of attention-related V1 activity in perceptual decision-making – and not only for processing visual stimuli – was also suggested (Silvanto et al., 2005). Report-related late activity is also thought to arise from recurrent feedback emitted by higher-order cortices (Pennartz et al., 2019), in agreement with the predominance of report-related coding in deeper cortical layers (Fig. 3.4c, see also Fig. 3.2f). In our experiments, silencing late V1 activity abolished the detection of orientation changes and contralateral stimuli in conditions of high cognitive demands (and consequently long reaction time), as is the case when having to process multiple sensory modalities simultaneously. This demonstrates a causal role of late V1 activity in perceptual decision making which is independent of scene complexity (cf. (Kirchberger et al., 2021)). However, this result is in apparent conflict with recent studies, which showed that late inactivation of V1 did not affect perceptual decisions (Resulaj et al., 2018), unless stimuli to be processed became more complex (Kirchberger et al., 2021). A possible explanation lies in the fact that, for relatively simple tasks, the window of temporal involvement of V1 in perceptual decision making might overlap with early, sensory-evoked activity. In line with this, the effectiveness of late inactivation was most prominent for low-saliency stimuli in MST mice and linearly scaled with reaction time (Fig. 3.5j, Fig. 3.6e). The only exception was the detection of high contrast stimuli in MST (but also UST) mice in task B, which was affected neither by early nor late V1 silencing (Fig. 3.6d, e). This suggests that subcortical structures (e.g. superior colliculus) may suffice to localize highly salient stimuli (Krauzlis et al., 2018) in task B (in contrast to task A, which requires detecting an orientation change), although we cannot fully exclude that portions of V1 were not completely inactivated. Furthermore, it is unlikely that late V1 silencing generally impaired cortical network processing (Otchy et al., 2015), as it did not affect ipsilateral visual detection in the visuotactile detection task (task B; Fig. 3.6), nor tactile or auditory performance (Supplementary Fig. 3.6b, Supplementary Fig. 3.8d). A final point relates to the increase in FAs following inactivation of late V1 activity in Task A (Fig. 3.5f, Supplementary Fig. 3.6f), which might suggest a different pathway for the observed behavioral effects. First, the criterion parameter of the psychometric model indicated that the behavioral effect of late V1 inactivation was not driven by this increase in FAs. Second, no increase in FAs was found in task B, which indicates that this increase may be specific to task A, in which mice were trained to react to any change in visual stimuli. Indeed, inactivation of V1 might be interpreted as a change in visual scene or view, as indicated by the fact that V1 inactivation did not increase FAs during auditory trials; moreover, no increase in FAs was observed in the control inactivation of S1 (Supplementary Fig. 3.7).

We optogenetically silenced both sensory and report-related components of V1 activity, which are jointly present in the late window (Fig. 3.3c). Importantly, late V1 inactivation in the absence of an early sensory-related component (e.g. ipsilateral visual stimuli in task B, or non-visual hit trials) did not impair behavioral responses, in agreement with a recent study suggesting that the report-related component alone is not sufficient for perceptual task performance (Kirchberger et al., 2021; Zatka-Haas et al., 2021). However, the question remains how sensory-evoked and report-related activity are related to each other during the late phase (Zatka-Haas et al., 2021). On the one hand, increased task demands may

prolong sensory processing of the visual stimulus, or at least the time downstream regions need to sample the ongoing flow of V1 activity to gather sufficient evidence on visual stimulus change (Brody and Hanks, 2016; Hanks et al., 2015). On the other hand, recurrent interactions between visual cortex and connected regions during late windows may jointly influence sensory representation, in line with the predictive processing framework (Leinweber et al., 2017; Pennartz et al., 2019; Rao and Ballard, 1999). In other words, in conditions requiring extended processing time, V1 and downstream regions need to interact for longer periods to jointly construct a behaviorally conclusive representation of the modality-specific change (task A) or the side of stimulus presentation (task B). The co-occurring increase in firing rate and drop in NCs that precede reaction time in hit trials likely originates in the pre-motor cortex, as previously suggested (Allen et al., 2017b). However, the question remains whether the behavioral effect of late V1 inactivation is due to a prolonged window during which downstream regions process information coming from V1, or rather if the late-onset report-related activity and the related drop in NCs recursively interact with ongoing sensory processing (Leinweber et al., 2017; Schneider and Mooney, 2018; Zagha et al., 2013) to shape perceptual decision making. Further experiments will be required to fully understand the mechanisms linking late V1 activity to perceptual report.

In conclusion, our results show that, although all sensory information that is theoretically required to perform a task is available in V1 shortly after stimulus onset, transforming such sensory inputs into a perceptual representation requires substantial recurrent interplay between cortical areas, which is temporally extended by factors increasing task demands (such as multisensory interactions). Our results thus dispute the classical picture of perceptual decision making: late-onset activity in primary visual cortex, which primarily stems from cortico-cortical recurrent interactions (Kirchberger et al., 2021; Li et al., 2013; Lien and Scanziani, 2013), is not simply involved in relaying, refining, and modulating the processing of complex visual stimuli, but also provides a causally relevant temporal window for perceptual decision-making.

Methods

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Umberto Olcese (u.olcese@uva.nl).

Materials availability

The study did not generate any unique reagents.

Experimental subjects

All animal experiments were performed according to national and institutional regulations. The experimental protocol was approved by the Dutch Commission for Animal Experiments and by the Animal Welfare Body of the University of Amsterdam. We used two transgenic mouse lines: PVcre (B6;129P2-Pvalbtm1(cre)Arbr/J, RRID: IMSR_JAX:008069) and F1 offspring of this PVcre line and Ai9-TdTomato cre reporter mice (Gt(ROSA)26Sor$^{tm9(CAG-tdTomato)Hze}$ RRID: ISMR_JAX 007909). A total of 49 male mice were used for this study:

• Task A: 28 mice (NE group: 7 mice, UST group: 4 mice, MST group: 17 mice)

• Task B: 12 mice (UST group: 4 mice, MST group: 8 mice)

Littermates were always assigned to the same experimental group. Mice in our colony are backcrossed to C57BL/6J wild-type mice from Jackson Laboratories every 10 generations. Mice were at least 8 weeks of age at the start of experiments. Mice were group-housed in a pathogen-free facility under a reversed day-night schedule (lights were switched off at 8:00 and back on at 20:00). All experimental procedures were performed during the dark period. Temperature in the housing facility was maintained between 19.5 and 23.5 C, and humidity was kept in a range between 45 and 65%. This study did not involve randomization or blinding. We did not predetermine the sample size. Some subjects were unable to successfully learn to make decisions based on both modalities (MST task versions) within 2 months and were excluded from further experiments. Although experiments were performed in male mice only, recent studies suggest that no difference should be expected in female mice as concerns the neuron-level mechanisms of perceptual decision making – see e.g. (Kirchberger et al., 2021; Steinmetz et al., 2019).

Head-bar surgery

Before the start of any experiments, mice were implanted with a headbar to allow head-fixation. Mice were subcutaneously injected with the analgesic buprenorphine (0.025 mg/kg) and maintained under isoflurane anesthesia (induction at 3%, maintenance at 1.5–2%) during surgery. The skull was exposed and one of two types of custom-made titanium circular head-bars with a recording chamber (version 1: inner diameter 5 mm, version 2: inner diameter 10 mm) was positioned over the exposed skull to include V1 and attached using C&B Super-Bond (Sun Medical, Japan) and dental cement. For task A in which visual stimuli were centrally presented binocular V1 was targeted based on the following coordinates (relative to lambda): AP 0.0, ML +/- 3.0 (Fong et al., 2016). Whereas coordinates sufficed for task A, for task B, in which lateralized visual stimuli were used, V1 was targeted using intrinsic optical imaging (see below) to localize the retinotopic region of V1 corresponding to the region of visual space in which the lateralized visual stimuli were presented. The skin surrounding the implant was covered using tissue adhesive (3M Vetbond, Maplewood, MN, United States) to prevent post-surgical infections. The recording chamber was covered with silicon elastomer (Picodent Twinsil). Mice were allowed to recover for 2-7 days after implantation, then habituated to handling and head-fixation before starting on the training procedure.

Behavioral training

Mice were subjected to a water restriction schedule and minimum weight was kept above 85% of their average weight between P60-P90. They typically earned their daily ration of liquid by performing the behavioral task but received a supplement when the earned amount was below a minimum of 0.025 ml/g body weight per 24h. Mice received ad libitum food.

Mice were head-fixed in a custom-built headbar holder in a dark and sound-attenuated cabinet. The body of the mouse was put in a small tube to limit body movements. The task was controlled in Octave (GNU Octave 4.x) interfacing with Arduino microcontroller boards (Arduino Uno, with code compiled in Arduino IDE 1.0.8). Licks were detected by capacitance-based or piezo-electric-based detectors. Upon correct licking, i.e. in hit trials,

5-8 µl of liquid reward (infant formula) was delivered immediately using gravitational force and solenoid pinch valves (Biochem Fluidics). Reward volume was calibrated biweekly to prevent lateralized response bias due to unequal reward size.

Behavioral task A: Audiovisual Change Detection

Auditory and visual stimuli were continuously presented throughout a behavioral session. During visual trials a feature changed (the orientation of a drifting grating), after which the visual display of this altered feature continued (post-change orientation) until the next visual change, and similarly for the auditory stimuli (post-change frequency).

Visual Stimuli

Visual stimuli consisted of full-field drifting square-wave gratings that were continuously presented with a 60 Hz refresh rate on an 18.5-inch monitor positioned at a straight angle with the body axis from the mouse at 21 cm from the eyes. Gratings were presented with a temporal frequency of 1.5 Hz and spatial frequency of 0.08 cycles per degree at 70% contrast and gamma-corrected. In trials with a visual change the orientation of the drifting grating was instantaneously changed (e.g. from 90˚ to 120˚) while preserving the phase. The degree of orientation change determined the visual saliency and was varied across experimental conditions.

Auditory Stimuli

Each auditory stimulus was a combination of five pure tones at harmonic frequencies. It was composed mainly of a center tone, as well as two lower and two higher harmonics (octaves below and above the center tone). If f_0 is the center tone: $f_{-2} = 1/4 * f_0$, $f_{-1} = 1/2 * f_0$, $f_0 = f_0$; $f_{+1} = 2 * f_0$; $f_{+2} = 4 * f_0$. All frequencies were expressed in scientific pitch as powers of 2 with the center tones spanning from 2^{13} Hz (=8372 Hz) to 2^{14} Hz (=16744 Hz). An example stimulus, $2^{13.5}$ (named by center tone), was therefore composed of five pure tones of $2^{11.5}$, $2^{12.5}$, $2^{13.5}$, $2^{14.5}$, and $2^{15.5}$ Hz. The weight with which each tone was present was taken from a Gaussian distribution across all tones for all stimuli, centered at $2^{13.5}$ (=11585 Hz). Because of this fixed weight distribution, stimuli with higher center tone frequency have decreasing weights for higher harmonics and increasing weights for lower harmonics. Stimuli with higher center frequency are thus increasingly made up of lower frequency components to the point of arriving at the starting stimulus (see also Supplementary Figure 3.1). This auditory stimulus design with harmonics and fixed weights was inspired by the Shepard tone illusion (Shepard, 1964). However, in contrast to this illusion, our stimuli were static and not sweeping across frequencies, and the original illusory aspect of a tone ever-increasing (or decreasing) in pitch was not exploited. The primary reason for this auditory stimulus design was the circular nature of the stimulus set, which mirrored the visual stimulus set with drifting gratings in all orientations.

During auditory trials, one stimulus was changed instantaneously to another, resulting in a shift in spectral power to five new frequencies. Auditory changes were expressed as partial octaves, with ½ octave maximally salient and the minimal change used was 1/256 partial octave. The degree of frequency/octave change determined the auditory saliency and was varied across experimental conditions. During auditory stimulus changes, the phase across all tones was preserved. Stimuli were presented with a sampling rate of 192 kHz. Stimuli were high-pass filtered (Beyma F100, Crossover Frequency 5-7 kHz) and delivered through two bullet tweeters (300 Watt) directly below the screen. Note that this high-pass filter

eliminated the lowest frequency components of the Shepard stimuli, and left the mid and high frequency components intact (those that span the sensitive part of the mouse hearing range, 8-16 kHz). This was done to prevent damage to the specialized tweeters that we used, but did not affect the animals' ability to report even very small differences between subsequently presented Shepard tones. Sound pressure level was calibrated at the position of the mouse and volume was adjusted per mouse to the minimum volume that maximized performance (average ±70 dB).

In an earlier cohort of mice trained on task A (N=13/28), the Shepard tones (1) were expressed in absolute Hz (e.g. an auditory trial with Δ2kHz changed from 8 kHz to 10 kHz), (2) had 9 harmonics, (3) were presented with a sampling rate of 48 kHz and (4) were not phase-preserved during a change in auditory frequency. We observed no qualitative or quantitative differences in both neural and behavioral results between the cohorts (behavior between cohorts is compared in Supplementary Fig. 3.2). The horizontal axes were normalized in Fig. 3.1 to accommodate all mice.

With both auditory and visual stimulus sets being circular, the direction of change (clockwise or anticlockwise) was behaviorally irrelevant (isotropy), and the only relevant dimension was the amount of change. Given the use of the full auditory spectrum and full-field visual gratings, stimuli in both modalities allowed change detection based on feature selectivity while recruiting neurons across the tonotopic organization of auditory cortex(Issa et al., 2014) and across the retinotopic map of visual cortex - which in our case benefitted both neural data acquisition and interventions.

Versions of Task A

Animals were assigned to one of three versions of a change detection task (NE, UST, MST) in which the visual and auditory stimuli were identical and only the reward contingencies varied. As we performed additional experiments with animals from the MST cohort, this resulted in a higher number of animals in the MST cohort.

NE: Noncontingent exposure (N=7/28 animals) – In this version, neither modality was associated with reward availability. Both the auditory and visual stimuli were continuously presented with the same distribution of trial types and temporal statistics as the multisensory version (see below). To compare intermittent licks, rewards, and stimuli across task versions, we sought to achieve similar rates of licking and reward delivery. Therefore, mice in this version could obtain rewards in a hidden 'response window' (a 1500 ms time interval in which either left or right licks could be emitted to acquire reward; same duration as MST, below). This response window was temporally decorrelated from the stimuli. Mice thus licked spontaneously at the two spouts and received occasional rewards. Mice were exposed 2-5 days to this behavioral task before any experiments.

UST: Unisensory version (N=4/28 animals) – In this version, only visual change was associated with reward availability. Mice were trained to respond to the visual changes only. Continuous auditory stimuli and changes were presented throughout training and recording sessions with the same statistics as the multisensory version, but were not associated with reward and were temporally decorrelated from the task-relevant visual trials. Given that only one side was rewarded in this version, spontaneous licking to this side had a higher probability of being rewarded and therefore the response window was shortened to 1000 ms (i.e., in this window, licks could be produced to acquire reward).

MST: Multisensory version (N=17/28 animals) – In this version visual and auditory change were both associated with reward availability. Mice were trained to respond in a lateralized manner to each modality: lick to one side to report visual changes, to the other side in case of auditory changes (modality-side pairing was counterbalanced across mice). Therefore, in this version, subjects had to simultaneously monitor both the auditory and visual modality, detect changes in a feature and discriminate the modality in which the change occurred. In other words, mice were required to identify the sensory modality in which a change occurred.

Training stages

For each trained modality (vision in UST, vision and audition in MST), training occurred in steps. In the first stage learning was facilitated by (1) only including the easiest trial type (maximally salient trials: 90 degrees orientation change for the visual domain(Goltstein et al., 2013) and 4kHz or ½ octave – in earlier and later cohorts, respectively – for the auditory domain), (2) additional instantaneous changes to increase saliency, (3) a passive reward on the correct side if the animal did not respond within 900 ms, and (4) the opportunity to correct after choosing the incorrect side. These facilitating conditions were phased out throughout the training procedure and trials of varying lower saliency were introduced. Animals were trained until their psychometric curve in the target modalities reached a plateau. For the MST version, animals were first trained in one modality, then the other, after which they were combined (the order of modalities was counterbalanced across mice).

Trials types were pseudorandomly presented (block-shuffled per 10 trials, 10% of trials were catch trials, thus without a stimulus change, 41% visual trials, 41% auditory trials, 8% multimodal trials - see below). The inter-trial interval was taken randomly from an exponential distribution with a mean of six seconds (minimum 3 and maximum 20 seconds). Directly after a stimulus change, a response window of 1500 ms followed in which either left or right licks could be emitted to acquire a reward. Licks during the first 100 ms were not counted as these occurred too early to be considered part of a stimulus-response sequence. The first lick after this 'grace period' was registered as the animal's choice and correct licks were directly rewarded. To counter any bias in MST mice, if the fraction of licks to one spout out of all licks in the last 10 trials was above 90%, the next trial was selected with a 95% probability to be of the other modality. As visual and auditory feature changes were associated with conflicting motor actions (only in the multisensory version of the task), a multimodal trial (simultaneous audiovisual change) would present the animal with conflicting signals. We introduced these conflict trials in a subset of sessions, but these trials were not included in the current analyses.

For each trained animal (before any recordings) we implemented three behavioral sessions in which we presented five levels of auditory and visual saliency that spanned the perceptual range to establish the perceptual sensitivity of each mouse. We fit the concatenated data of these three sessions with a cumulative normal distribution per modality with four free parameters (Meijer et al., 2018):

$$f(x) = \gamma + (1 - \gamma - \lambda)\left(\frac{1}{2}\left[1 + \operatorname{erf}\frac{x - \mu}{\sigma\sqrt{2}}\right]\right) \qquad \text{(Eq. 3.1)}$$

Here, γ describes the false alarm rate (spontaneous licks during catch trials), λ the lapse rate (misses at maximal saliency), μ the mean (perceptual threshold), and σ the standard deviation (sensitivity to variations of stimulus intensity). Having established the psychometric function per mouse, we took four levels of saliency per modality at fixed points along the psychometric function: subthreshold (μ-σ; *sub*), threshold (μ; *thr*), suprathreshold (μ+σ; *sup*), and maximal saliency (*max*). The visual threshold ranged from 4-12 degrees, and the auditory threshold from 10-100 Hz (frequency version) or 1/64 - 1/16 partial octave (octave version) (Supplementary Fig. 3.2). This analysis was purely performed to select stimulus intensities of equal subjective saliency across mice for the experiments. All other analyses were based on fitting the behavioral data with a psychometric signal detection model (see below).

In recording sessions, we limited conditions to sample sufficient trials per modality x feature x saliency x choice combination. First, we only used two levels of change: threshold and maximal saliency. For NE mice and auditory conditions in UST mice, we used threshold values that matched those from trained animals. Second, we only used four orientations or tones. Specifically, this means that stimuli jumped between A, B, C, and D, where distance AB and CD are around threshold and distance AC and BD are maximal. An example stimulus set for a mouse with a visual threshold of 7° and an auditory threshold of 1/32 octave was therefore for the visual domain: A=100°, B=107°, C=190°, D=197°, and for the auditory domain (in Hz): $A=2^{13.25}$, $B=2^{13.25+1/32}$, $C=2^{13.75}$, $D=2^{13.75+1/32}$.

Behavioral task B: Visuotactile Side Detection

As in paradigm A, mice were trained in on one of two versions of a visuotactile detection task: a multisensory version, where both visual and tactile modalities were informative on the side that needed to be chosen to acquire reward (MST) and a unisensory version, where the tactile modality was present as well, but only the visual modality was informative (UST).

Stimuli used

Visual Stimuli

Visual stimuli consisted of square-wave drifting gratings, with a temporal frequency of 1.5 Hz, a spatial frequency of 0.025 cycles per degree and 30 degrees orientation. The contrast of the gratings was modulated per trial to control detection difficulty. Visual stimuli were generated in Octave using Psychtoolbox3 and were presented monocularly at >24 degrees from azimuth on each side(Heesy, 2004) with a gamma-corrected 18.5-inch monitor at a frame rate of 60 Hz and a distance to the eye of 18 cm.

Tactile Stimuli

Tactile stimuli consisted of a single deflection of the whisker pad using a piezoelectric bender (PL128.10, Physik Instrumente) coupled to a 5 cm long pipette ending on a 5x5 mm patch of Velcro. A voltage driver (E650, Physik Instrumente) and an RC filter were used to produce a backward deflection of the bender with an exponentially decaying speed (τ=72 ms) during 360 ms, followed by a forward deflection with the same characteristics. The amplitude of the deflection was modulated to control detection difficulty. Elicited whisker deflection angles ranged from 0 to 3.6 degrees. For both visual and tactile stimuli, stimulus intensity was adjusted individually to match the desired saliency.

Versions of task B:

UST: Unisensory version

Visual and/or tactile stimuli were presented to either the right or left side of the animal. To obtain a reward, mice had to detect the side where the visual stimulus was presented and lick the spout at the corresponding side. In this version, only the visual modality was informative on reward availability. Tactile stimuli were delivered but not associated with reward and tactile and visual stimulus sides were decorrelated. The inter-trial interval was drawn from an exponential probability distribution with a mean of 4 seconds (minimum 3, maximum 7; with a 22% chance of catch trial (no stimulus, no reward) and a maximum of two catch trials in a row, a mouse could wait up to 21 seconds before another stimulus was displayed). Visual and/or tactile stimuli were presented for 1 second. In a multisensory trial (not analyzed here), the tactile stimulus was presented with a lag of 70 ms after the visual stimulus onset (similar to (Olcese et al., 2013)). Licks were only rewarded in the interval of 140-1000 ms after stimulus onset. While a correct lick triggered reward delivery, an incorrect lick (i.e., to the wrong side) terminated the trial and aborted stimulus presentation. Trials were pseudo-randomly generated by blocks of 60 with 22% catch trials, 12% tactile-only trials, 53% visual-only trials, and 13% multisensory trials.

MST: Multisensory version

In this version, both visual and tactile modalities were informative on reward availability. In multisensory trials, visual and tactile stimuli were presented on the same side. Overall, task B required the mouse to follow an Inclusive-Or rule (lick to the side with either a visual or tactile stimulus, or a compound stimulus in both modalities). During training, mice first learned to detect tactile stimuli. Multisensory trials were then added and finally, visual-only trials were introduced so that mice could eventually detect visual and/or tactile modalities. Since tactile trials were rewarded, to keep the reward/no-reward balance, we increased the number of tactile trials: 25% catch, 25% visual-only, 25% tactile-only, 25% multisensory. Otherwise, both unisensory and multisensory task versions had the same parameters.

The noncontingent exposure (NE) version was not implemented for task B.

Imaging, Optogenetics, and Electrophysiology

Intrinsic Optical Imaging

To localize the primary visual cortex in task B experiments, we performed intrinsic optical imaging (IOI) under lightly anesthetized conditions (0.7-1.2% isoflurane). A vasculature image was acquired under white light before starting the imaging session. During IOI, the cortex was illuminated with 630 nm light and images were acquired using a CCD camera connected with a frame grabber (Imager 3001, Optical Imaging Inc, Germantown, NY, USA), defocused about 500 μm below the pial surface. Visual stimulation consisted of square-wave drifting gratings (duration 8 s, 2 Hz, 0.05 cycles/deg, 100% contrast) presented in the right visual hemifield. All image frames obtained during stimulus presentations were divided by the average of the first 10 frames acquired just before stimulus presentation. The target area was outlined as the region with visually-evoked decrease in reflectance, using custom-made software in MATLAB (Olcese et al., 2013).

Viral injection

We performed viral injections and optogenetic experiments in a total of 25 mice, representing a subset of the full experimental cohort (Task A: n=4/4 UST, n=9/17 MST; Task B: n=4/4 UST, n=4/8 MST). Mice were subcutaneously injected with the analgesic buprenorphine (0.025 mg/kg) and maintained under isoflurane anesthesia (induction at 3%, maintenance at 1.5–2%) during surgery. We performed small craniotomies (±100 µm) over V1 using an ultrafine dental drill and inserted a glass pipette backfilled with AAV2.1-EF1a-double floxed-hChR2(H134R)-EYFP-WPRE-HGHpA (titer: 7×10^{12} vg/mL, 20298-AAV1 Addgene). In total 50 nL was injected in V1 (bilateral binocular V1 for Task A and unilateral V1 for Task B) at 700 µm and 400 µm below the dura (25 nL per depth) using a Nanoject pressure injection system (Drummond Scientific Company, USA).

Optogenetics

In a random subset of trials (50% of trials for task A, 25% for task B) photostimulation started at stimulus onset (early inactivation) or was delayed (late inactivation). For the MST version of task B, early and late inactivation took place in separate sessions. Late inactivation occurred after 200 ms in Task A and 240 ms in Task B. Photostimulation continued until the animal made a choice. We interleaved sessions in which we positioned the fiber over V1 with control sessions in which we either positioned the optic fiber over area S1 (where no virus was injected) or at the head-bar. To locally photostimulate V1, a 473 nm laser (Eksma Optics, DPSS 473nm H300) was connected to one or two fiber-optic cannulas (ID 200 um, NA 0.48, DORIC lenses) that were positioned directly over the thinned skull at the area of interest (bilateral V1 for Task A and unilateral V1 in Task B). Light delivery was controlled by a shutter (Vincent Associates LS6 Uniblitz) with variable pulse and interpulse duration with an average of 20 Hz and 75% duty cycle (Task A) or with 10 ms pulses sequentially interleaved by 20 ms and 30 ms (~72% duty cycle, Task B). The shutter was located in a sound-insulated box distal from the experimental setup. As we simultaneously performed extracellular recordings in V1 of all mice, we adjusted laser power for each animal to the minimum power that maximally inhibited neural activity. This was commonly 2-15 mW at the tip of the each fiber (placed 0.5-2 mm above the cortical surface) corresponding to an effective 1.5-5.25 mW/mm^2 at the cortical surface (taking the ~75% duty cycle into account), which is below the levels that produce unwanted heating in tissue (Kirchberger et al., 2021; Stujenske et al., 2015).

To prevent light from reaching the eye of the mouse, the cannulae were sealed with black tape, leaving only the tip exposed. Furthermore, sessions with optogenetic manipulation were performed in an environment with ambient blue light. Even though we implemented these measures, we observed an increase in false alarms in some mice in task A. This suggests either that mice could perceive the laser, or that our manipulation evoked perceptual changes that were reported as a trial. We therefore verified (1) that our main effect of late silencing was not explained by a change in criterion (see Behavioral Analysis Task A), (2) positioned the fiber over uninfected somatosensory cortex (S1), and (3) performed the same optogenetic experiments in a second visuotactile paradigm where we did not have an increase in False Alarm responses by photoinactivation of V1. In Task B, due to a lack of sufficient trials to test optogenetic silencing for all conditions, only the reported conditions were tested (Visual Contralateral Thr and Max, Visual Ipsilateral Thr, Tactile Contralateral Thr) as well as Multisensory Contralateral Thr.

Extracellular recordings

Mice were subcutaneously injected with the analgesic buprenorphine (0.025 mg/kg) and maintained under isoflurane anesthesia (induction at 3%, maintenance at 1.5–2%) during surgery. We performed small (about 200 μm) craniotomies over the areas of interest (up to 6 per animal) using a dental drill. The recording chamber was sealed off with silicon elastomer and the mice were allowed to recover for 24h.

Extracellular recordings were performed on consecutive days with a maximum of 4 days to minimize damage to the cortex. Microelectrode silicon probes (NeuroNexus, Ann Arbor, MI – 4 types of either 32 or 64 channels were used, catalog numbers A1x32-Poly2-10mm-50s-177, A2x16-10mm-100-500-177, A4x8-5mm-100-200-177, A1x64-Poly2-6mm-23s-160) were slowly inserted in the cortex until all recording sites were in contact with the tissue. V1 was approached perpendicularly to the cortical surface. The medial prefrontal cortex, primary auditory cortex, and posterior parietal cortex were also recorded, but data from these areas were not analyzed here. After insertion, the exposed cortex and skull were covered with 1.3-1.5% agarose in artificial CSF (125 mM NaCl, 5 mM KCl, 1.3 mM $MgSO_4$, 2.0 mM NaH_2PO_4, 2.5 mM $CaCl_2$, pH 7.3) to prevent drying and to help maintain mechanical stability. The probe was left in place for at least 15 minutes before recording to allow for tissue stabilization. Electrodes were dipped in DiI (ThermoFisher Scientific) during the final recording session allowing better post hoc visualization of the electrode tract. The ground was connected to the head bar and the reference electrode to the agarose solution. Neurophysiological signals were pre-amplified, bandpass filtered (0.1 Hz to 9 kHz), and acquired continuously at 32 kHz with a Digital Lynx SX 64/128 channel system in combination with the acquisition software Cheetah 5.0 (Neuralynx, Bozeman, MT).

Spike sorting of data acquired during task B was done with custom-made software in MATLAB, as previously described (Olcese et al., 2013), and only units having less than 1% of their spikes within a 1.5 ms refractory period were kept. For task A we used Klusta 3.0.16 and then manually curated with the Phy GUI (Rossant et al., 2016) (Phy 1.0.9). Before spike sorting the median of the raw trace of nearby channels (within 400 μm) was subtracted to remove common artifacts. Each candidate single unit was inspected during manual curation based on its waveform, autocorrelation function, and its firing pattern across channels and time. Only high-quality single units were included, defined as having (1) an isolation distance higher than 10 (cf. (Schmitzer-Torbert et al., 2005)) (2) less than 0.1% of their spikes within the refractory period of 1.5 ms (Bos et al., 2017; Vinck et al., 2015b), (3) spiking present throughout the session. Neurons were deemed stably present if they had spikes in more than 90 out of 100 time bins during the entire session.

Recording depth estimation

The estimation of the laminar depth of the electrodes in V1 was based on three aspects. First, we computed the power in the 500-5000 Hz range to localize layer 5 with the highest MUA spiking power (Senzai et al., 2019). Second, we showed contrast-reversing checkerboards before each recording session and computed the current source density profile to estimate layer 4 with the earliest current sink, as previously described (Chapter 2). Lastly, this was aligned with the depth registered when the silicon probes were lowered from the dura. The granular layer was taken to span from 400 to 550 μm from the dura.

Video monitoring

In Task A, the left eye (ipsilateral to the hemisphere of recording) was illuminated with an off-axis infrared light source (six infrared LEDs 850 nm) adjusted in intensity and position

to yield high contrast illumination of both the eye and whisker pad. A frame-grabber acquired images of 752x582 pixels at 25 frames per second through a near-infrared monochrome camera (CV-A50 IR, JAI) coupled with a zoom lens (Navitar 50 mm F/2.8 2/3" 10MP) that was positioned at approximately 30 centimeters from the mouse.

To extract pupil variables (Meijer et al., 2020; Montijn et al., 2015) we trained DeepLabCut(Mathis et al., 2018) (version 2.1.10) on 300 frames from 15 video excerpts of 1-2 minutes with varying pupil size, illumination, contrast, imaging angle, and task conditions. We labeled the pupil center and 6 radially symmetric points on the edge of the pupil. An ellipsoid was fit to these 6 outer points. The pupil center was taken as the center of the ellipsoid and the pupil area as the ellipsoid area from the fitted ellipse parameters. Single poorly fit frames were replaced by the running median (10 frames). We z-scored the total session trace.

Histology

At the end of each experiment, mice were overdosed with pentobarbital and perfused (4% paraformaldehyde in phosphate-buffered saline), and their brains were recovered for histology to verify viral expression and placement of silicon probes in V1. We cut coronal 50 µm sections with a vibratome, stained them with DAPI, and imaged the mounted sections. For flattened cortical sections (e.g. Fig. 3.5b) we first removed subcortical tissue and flattened the cortical sheet of each hemisphere between glass slides by applying pressure overnight before sectioning 100 µm slices with the vibratome (Lauer et al., 2018). For coronal sections, area borders were drawn by aligning and overlaying the reference section from the atlas(Paxinos and Franklin, 2004). For flattened cortical sections, area borders were drawn based on cell densities aligned to reference maps (Gămănuţ et al., 2018).

We frequently observed a minor reduction in fluorescence at the center of the viral injection site. Therefore, a potential concern could be that the cortical circuit at the center of the site could be either affected in function or less responsive to optogenetic inactivation. However, visual task performance was similar before and after viral injection (d-prime around 1.5 in maximal visual saliency control trials; Fig. 3.1f and 3.5g). Moreover, optogenetic inactivation was highly effective in silencing pyramidal cell activity and in impairing visual discrimination for difficult visual trial types during both behavioral tasks (Fig. 3.5g; 3.6d).

Quantification and Statistical Analysis

Data analysis

Unless otherwise stated, all data were analyzed using custom-made software written in MATLAB (R2016a, The MathWorks, Natick, MA).

Behavioral analysis - Task A

Sessions were terminated when the animal did not respond for 20 trials and these last 20 trials were discarded from analyses. Sessions in which the hit rate for maximal auditory and visual changes was below 30% were excluded.

Behavioral response rates in task A were fit with a multi-alternative signal detection model (Sridharan et al., 2014). This model extends signal detection theory (Green and Swets, 1966) and aims to accurately and parsimoniously account for observer behavior in a

detection task with multiple signals. In this model, the decision is based on a bivariate decision variable whose components encode sensory evidence in each modality. Decision space is partitioned into three regions (no response: neither evidence is strong enough; auditory response, and visual response). In a given trial, the observer chooses to report visual or auditory stimuli if the decision variable exceeds a particular cutoff value, the "criterion" for each signal (the animal's internal signal threshold for responding, in terms of signal detection framework). We fit two versions of this model. In sessions with two levels of saliency (threshold and maximum), we fit the d-prime (d') and criterion (c) to the behavioral response rates separately for each stimulus change intensity. This consists of fitting four free parameters (d' and c for each modality). In sessions with four or five levels of saliency per modality, we fit the behavioral response rates by fitting a criterion per modality and a d-prime for each saliency, which is described by a psychophysical function (three-parameter hyperbolic function). The d-prime at each saliency level follows from:

$$d_i = d_m \quad x_i^n / (x_i^n + s_5^n) \qquad \text{(Equation 3.2)}$$

where d_{max} is the asymptotic d-prime, s_{50} is the stimulus strength at 50% of the asymptotic value, n is the slope of the psychometric function and x_i is the amount of change. This consisted of fitting a total of 8 free parameters: d_{max}, n, s_{50}, and c for each modality. We refer the reader to Sridharan *et al.* (2014) for a detailed description of how the d-prime and criterion subsequently relate to response rates. Single session fits where visual threshold was below 1 degree or above 45 degrees were excluded (average threshold ±6 degrees, n=3/179 sessions excluded).

Catch trials during the tasks (in all cohorts) served to measure baseline lick responses. As there were no stimulus changes during the inter-trial interval (visual and auditory stimuli continued to be presented throughout the session, similarly to catch trials), we used long inter-trial intervals to insert additional artificial catch trials during offline analysis to improve the statistical balance across trial type conditions. We controlled for the temporal expectation of stimulus change and inserted additional catch trials only at time points conforming to the original inter-trial interval statistics. These additional catch trials included both false alarms (i.e. spontaneous licks during inter-trial intervals) and correct rejections (no spontaneous licks). We further constrained the timing of these false alarms such that the distribution of their response latencies matched those of hit trials. For sessions with high false alarm rates, FA trials were subsampled to match the distribution of hit trials. Note that these additional catch trials only served for analysis purposes and not for measuring behavioral performance.

After analyzing the effects of early and late V1 silencing on audiovisual change detection on the full dataset, we focused in a subsequent analysis on the relationship between reaction time and the effect of late silencing (Fig. 3.5i, j, Supplementary Fig. 3.6d-f). Here, we focused on sessions in which V1 early silencing was effective (minimum 50% reduction in d-prime on maximal visual change, 59/81 sessions; results were robust to variations of this criterion, 25% reduction, r=0.199, p=0.049; 75% reduction, r=0.511, p=0.001). This threshold was implemented to test if late silencing was effective specifically within those sessions in which the optogenetic manipulation demonstrably impaired visual detection (thus exploiting an internal control).

Behavioral analysis - Task B

Behavioral data in task B was fit with a multinomial logistic regression(Burgess et al., 2017). The probabilities of right choice (p_{right}), left choice (p_{left}) and no choice ($p_{no\text{-}go}$) were set by:

$$\log\left(\frac{p_{r\ \ nt}}{p_{n\ -g}}\right) = b_{r\ \ nt} + s_{.\ r\ \ nt} \square c^{\ n} + s_{.\ r\ \ nt} \square c_{.}^{\ n} \quad (3.3)$$

$$\log\left(\frac{p_{l\cdot}}{p_{n\ -g}}\right) = b_{l\cdot\ \ } + s_{.\ l\cdot\ \ } \square c^{\ n} + s_{.\ l\cdot\ \ } \square c_{.}^{\ n} \quad (3.4)$$

Where b is a bias parameter, sL and sR are the sensitivity to stimulus evidence to the left and right side respectively, cL and cR are the stimulus intensity to the left and right side respectively (contrast for vision, deflection angle for somatosensation), n is an exponent parameter between ranging between 0 and 1 to allow for saturation.

The model was fit to individual mice, with all sessions pooled together. However, per mouse, visual behavior (visual-only trials) and tactile behavior (tactile-only trials) were fit separately. The model was fit using Matlab's *mnrfit* and maximum likelihood estimation. To quantify behavioral performance, we computed d-prime (d') as:

$$d' = \phi'(\%C\quad r_{\cdot}\quad t_{\cdot}\ s_{\cdot}\quad i) - \phi'(\%F\quad a_{\cdot}\quad t_{\cdot}\ s_{\cdot}\quad i) \quad (3.5)$$

Where ϕ' is the normal inverse cumulative distribution function.

Electrophysiological data processing

To visualize the effect of photostimulation on spiking activity at a single electrode channel the raw signal was high-pass filtered (500 Hz, 4th order Butterworth filter). To compute firing rates, spikes (following spike detection and sorting) were binned in 10 ms bins and convolved with a causal Half-Gaussian window with 50 ms standard deviation, unless stated otherwise. Wherever firing rate was z-scored, the mean was subtracted and divided by the standard deviation of the baseline period (-1 to -0.2 seconds before stimulus). For Fig. 3.5d the firing rate was only normalized to the baseline to quantify the relative reduction in firing rate by optogenetic inhibition. For Fig. 3.7g and 3.7h the standard deviation of the convolutional window was reduced to 10 ms to enhance temporal resolution. For computing noise correlations the standard deviation of the convolutional window was increased to 100 ms to increase noise correlation estimates. For Fig. 3.6c and Supplementary Fig. 3.8g, h, neurons with an average z-scored firing rate that exceeded 2 standard deviations at any point during the stimulus epoch of any visual trial condition were considered responsive and included. If neuronal activity was sampled in less than 3 trials in the relevant conditions they were excluded. Neurons in sessions lacking any of the compared conditions were excluded.

Encoding model of single neuron firing rates

To quantify single neuron encoding of different task variables, we constructed a kernel-based Poisson regression model. This encoding model allowed us to model, for single neurons, the time-dependent effects of all measured variables related to the task and the animal's behavior simultaneously on single-trial neuronal activity (Park et al., 2014; Runyan et al., 2017). This approach is particularly useful to disentangle the unique

contribution of experimenter-controlled task events and self-timed behavioral events to variability in firing rates across the neuronal population.

Construction - For each neuron, we constructed a design matrix based on five sets of variables; visual, auditory, hit, movement, and arousal variables. Binary variables (all except pupil size) were modeled with a series of temporal basis functions (raised cosines) that spanned the relevant epoch of influence. The number and temporal distribution of these basis functions were selected to maximize the cross-validated explained variance (see below). For the sensory predictors, we used two kernels with 100 ms standard deviation that spanned the first 200 ms post-stimulus to capture the early spiking activity and 10 kernels with 200 ms standard deviation that spanned from 0 to 2000 ms post-stimulus to capture the late, sustained response. We found that making a separate predictor set per combination of orientation x amount of change produced the highest quality fit as it simultaneously took into account the selectivity of neurons for orientation and saliency. This therefore resulted in (2 + 10 basis functions) x 2 (modalities) x 2 (levels of change) x 2 (grouped post-change features) = 96 predictors. For hit variables we used 10 temporal basis functions with 200 ms standard deviation that spanned from 0 to 2000 ms relative to stimulus change in hit trials (visual hit, audio hit) and 10 predictors that spanned -500 ms to +1500 ms relative to reward (20 predictors for hit). For movement variables, we used three basis functions that spanned -200 to +400ms relative to each lick, split by side (6 predictors). To capture arousal effects, the z-scored pupil area was included in the predictor set: with original timing and two temporal offsets (-800 ms and -400 ms) to account for the delayed relationship of brain state to pupil size (e.g. (McGinley et al., 2015b); this equals 3 predictors). We included one whole-trial variable that scaled with the within-session trial number. This full model summed up to 126 predictors. We compared the performance of this model to a null model, with one predictor (a random variable). For convenience, all predictors were normalized to their maximum values before being fed into the model.

Fitting - We fitted the encoding model to each neuron's activity individually, using the *glmnet* package in Matlab (Friedman et al., 2010) (2015 version) with elastic-net regularization and a Poisson link function, which involves setting three hyperparameters. First, we chose elastic net mixing parameter $\alpha = 0.95$ to allow for a small number of uncorrelated informative predictors to be favored. Second, model performance was trained and tested on separate data with 5-fold cross-validation. Third, to maximally punish weights without losing model fit quality, regularization parameter lambda was maximized while keeping the cross-validated error within one standard error of the minimum (lambda_1se in *glmnet*). Because very sparsely firing neurons produced fitting difficulties, only neurons with a session-average firing rate >0.5 Hz were included.

Evaluation - We quantified the model performance by assessing the 5-fold cross-validated Explained Variance (EV) by the predicted firing rate based on the random or full model, or a subset of predictors from the full model. Explained Variance was calculated as:

$$\mathrm{EV} = 1 - \frac{\mathrm{var}\left(Y - \hat{Y}\right)}{\mathrm{var}(Y)} \tag{3.6}$$

where Y is the original firing rate and $\hat{Y}$ the estimated firing rate. Explained Variance was computed in two ways. First, we computed EV over all concatenated firing rate bins (over all single trials; -0.5 to 2.5 seconds relative to stimulus change). Second, we computed EV

on the concatenated firing rate bins of the average firing rate for five trial-type x choice conditions that captured most trial counts(Musall et al., 2019; Runyan et al., 2017) (85% of all trials). To compute EV over time we computed the explained variance over all concatenated time bins at a specific moment relative to stimulus onset.

Decoding single neuron activity

To identify which variables could be decoded from single neurons we used ROC analysis (Green and Swets, 1966) and identified how well an external observer could discriminate variables from the firing rate at single time points. In contrast to the GLM encoding approach, this analysis is more suited to assess at which timepoint downstream neurons can discriminate task-relevant variables (e.g. by integrating spike rate over a small time window), in view of its superior temporal resolution due to the absence of kernel fitting. We computed the area under the ROC curve (AUC) for the firing rate distributions between two selections of trials. Each class had to have at least 10 trials. AUC values are in the range of 0 to 1 and were rectified to their corresponding values in the range between 0.5 and 1. We investigated three types of coding and for each of these we analyzed threshold change and maximum change trials separately:

Visual Orientation: We grouped the pairs of post-change orientations that were close to each other (e.g. A and B oriented at 90 and 97 degrees, see above) and thus compared the firing rate distributions of A&B versus C&D for threshold and maximal change trials separately.

Occurrence of visual change: We tested whether single neuron firing rates discriminated between visual and catch trials.

Hit/miss: To identify significant coding of the detection of a visual stimulus we compared firing rate distributions within visual trials for hits and misses.

To determine the significance of AUC values at each time bin and for each comparison, we performed a permutation test by shuffling the class labels across trials 1,000 times. If the unshuffled AUC value exceeded 99% of the shuffled distribution ($P < 0.01$) this was deemed significant. This yielded an AUC value for each neuron for each time bin for each type of coding and each of these values its significance by the permutation test. For the Venn diagram in Fig. 3.4A, neurons were classified as coding if their AUC score was significant for at least three consecutive time bins (=75 ms, permutation test) during the 0 to 1000 ms window after a stimulus change.

To compare coding dynamics across cohorts, we normalized the fraction of significantly coding neurons by subtracting the baseline fraction (average over -0.5 to 0 seconds) and dividing by the maximum. Each condition was only normalized to maximum if the fraction of significantly coding neurons increased at least 10% over baseline.

To determine the onset of significant coding we tested when the fraction of coding neurons increased significantly above a multiple of standard deviations of the coding fraction during baseline. We report results at a threshold of 2 standard deviations (Zscore > 2), but the results were robust to variations in threshold (e.g. 1 or 3 standard deviations). To estimate the reliability of the onset of coding and the relationship of hit/miss coding to reaction time, we bootstrapped by resampling from the total neuronal population (n=1000 bootstraps). To investigate the relationship between the onset of hit/miss coding and reaction time we

used a linear regression, which revealed a systematic relationship between the timing of hit/miss coding and reaction time (Fig. 3.4e, f). To estimate by how much hit/miss coding preceded reaction time we used two measures. First, we fixed the slope of the regression fit at 1 and found an offset of 278 ms. This was similar across variations of threshold (Z > 1: 288 ms, Z > 3: 250 ms). Second, for each bootstrap, we computed the onset of hit/miss-coding according to the fit parameters for the average reaction time, which was on average 266 ms before the reaction time.

Identifying the onset of coding by using the fraction of neurons coding above baseline was not suitable for individual sessions with low numbers of neurons (e.g. n=10 neurons). Therefore, to estimate the onset of hit/miss coding per session we performed the same AUC analysis, but now on the averaged firing rate across neurons for each session with at least 10 neurons. Similarly, per time bin the significance of hit/miss coding was tested against a shuffled distribution of trial labels (n=1000 shuffles). The onset was taken as the first significant time bin after stimulus onset.

For laminar depth localization of coding dynamics, neurons were binned according to their recorded depth in 50 µm bins spanning from 0 to 1150 µm below the dura. The fraction of neurons coding for each variable at this depth was computed for each time point (25 ms temporal bins). This heatmap was convolved for display purposes with a two-dimensional Gaussian (standard deviation of 1.3 bins – temporal and spatial). For statistical comparison across laminar zones, the fraction of coding neurons was computed for each session (if at least 10 neurons were recorded at this depth to estimate coding fraction reliably) in supragranular, granular, or infragranular layers (granular layer: 400-550 µm from dura). As sensory and hit/miss-coding was present in different temporal epochs these were included for statistical comparison (Orientation 0-1000 ms, Visual occurrence: 0-200 ms, Hit/miss: 200-1000 ms, relative to stimulus change).

Population coding analysis

To decode visual stimulus orientation, we departed from the four orientations and grouped the two pairs of orientations close to each other to obtain a two-class classification problem (AB vs CD, see above). Decoding was performed on recordings that contained at least 15 neurons and in which at least 20 trials per orientation pair were available. We equalized the number of neurons across sessions by randomly drawing 10 neurons from all sessions with more than 10 units. Spikes were binned using a sliding window of 200 ms with 50 ms increments, excluding time bins that contained both pre and post-stimulus spikes. Decoding was performed using a random forest classifier with 200 trees, as implemented in Scikit-learn (Pedregosa et al., 2011) (version 0.23.0), and we employed a 5x5 cross-validation routine with stratified folds (cf. (Bos et al., 2019)). The average accuracy obtained in the cross-validation routine was corrected by subtracting the average accuracy on 50 surrogate datasets in which the orientation labels were permuted across trials to obtain the improvement in decoding accuracy beyond chance level.

Noise Correlations

To investigate correlated activity across the population we computed pairwise correlations on the binned spike counts (10 ms bins, time range: -1000 to +1500 ms relative to stimulus change) after subtracting the average stimulus-driven response. First, for each neuron the trial-mean firing rate over time was subtracted for all subsets of trials of interest (per orientation). Next, the Pearson's correlation coefficient was computed between the residual

rates for each simultaneously recorded neuronal pair for each time bin. Neuronal pairs for a given condition were included if they were sampled in more than 10 trials. We also computed pairwise correlations aligned to lick onset and thus subtracted mean activity related to lick-related modulation of firing rate. We note, however, that the term 'noise correlations' is conventionally reserved for the correlations on residual rates after subtracting the mean stimulus-evoked activity (rather than movement-evoked activity). Note also that drop in noise correlations was not a direct result of overall increased firing rates as we observed no such reduction of noise correlations during early sensory-evoked activity (0 – 200 ms). Neurons were only included if their session average firing rate was above 1 Hz.

Statistics

Unless specified otherwise all statistics were performed using linear mixed models (LMMs) or generalized linear mixed models (GLMMs) in MatLab (MathWorks, Natick, MA). (G)LMMs take into account the hierarchical nature of our data (neurons and trials sampled from the same mice) (Aarts et al., 2014). (G)LMMs describe the relationships between a response variable and multiple explanatory variables, and comprise two types of explanatory terms. Fixed effects are the variables of interest, while random effects, also commonly referred to as grouping variables, specify and account for the nesting group (mouse ID in our case). For all analysis involving hierarchical data, LMMs were constructed with mouse ID as a random effect (intercept only). Importantly, mouse ID was not included as a random effect for analyses with cohort as fixed effect, as variability between mice was key to those results. As firing rates are generally not normally distributed, for analyses of firing rate responses (non z-scored) we used Generalized Linear Mixed Models with a Poisson distribution. Statistical tests were performed on the fixed effect using ANOVAs on the (G)LMMs. To estimate the denominator degrees of freedom (DF2) for F-tests, the Satterthwaite approximation was used for LMMs and the residual degrees of freedom for GLMMs. Linear hypothesis tests were performed in the case of post hoc comparisons using the relevant contrasts. A description of each test, sample sizes, test statistics, and *p* values are provided in Supplementary Table 3.1. Results with a p-value lower than 0.05 were considered significant.

To verify whether our cohort-specific results were likely to be a consequence of a difference in group size, we subsampled MST neurons and sessions to match UST sample size for two key results (Fig. 3.4d, 3.5h) and found that our results were maintained (1000 resamples, see Supplementary Table 3.1).

Data availability

All behavioral and neural data related to this study are openly available at https://gitlab.com/csnlab/olcese-lab/modid-project/2nd-bump and has been deposited on Zenodo (https://doi.org/10.5281/zenodo.6451263). Source data are provided with this paper.

Code availability

All code related to this study is openly available at https://gitlab.com/csnlab/olcese-lab/modid-project/2nd-bump. The version of the code used in this study is available on Zenodo (https://doi.org/10.5281/zenodo.6451263).

Acknowledgments

We thank D. Sridharan for providing code for the multi-alternative detection model; C. Rossant, members of the Cortex Lab (UCL) and contributors for Klusta and Phy spike sorting software; Andriana Mantzafou, Klara Gawor, and Alexis Cervàn Canton for assistance in behavioral training. This work was supported by the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement 720270 (Human Brain Project SGA1) to C.M.A.P., Grant Agreement 785907 (Human Brain Project SGA2) and 945539 (Human Brain Project SGA3) to C.M.A.P. and U.O., by the FLAG-ERA JTC 2015 project CANON (co-financed by the Netherlands Organization for Scientific Research – NWO) to U.O., by the FLAG-ERA JTC 2019 project DOMINO (co-financed by NWO) to U.O. and by the Amsterdam Brain and Mind Project to C.M.A.P. and C.P.K.

Author contributions statement

Conceptualization, M.O.L., J.P., C.M.A.P., U.O.; Methodology, M.O.L. and J.P.; Main analysis, M.O.L. and J.P., Additional analysis, P.M., J.M.; Data Curation, M.O.L. and J.P.; Writing – Original Draft, M.O.L., J.P, U.O.; Writing – Review & Editing, All; Visualization, M.O.L. and J.P.; Supervision, U.O., C.M.A., and C.P.K.; Funding Acquisition, U.O., C.M.A. and C.P.K.

Competing interests statement

The authors declare no competing interests.

Materials & correspondence

Correspondence and requests for materials should be addressed to U.O.

Supplementary Data

Supplementary Table 3.1 Statistical tests

Figure	Goal	Test	Sample sizes	Test statistic	p-value (significant in bold)
3.1f	Visual d-prime UST vs MST	Linear Mixed Model ANOVA (fixed effect: cohort)	12 UST sessions, 139 MST sessions	F(1,29) = 1.60, p=0.22	p=0.22
3.1f	Auditory vs visual d-prime in MST	Linear Mixed Model ANOVA (fixed effect: modality, random effect: mouse ID)	139 MST sessions	F(1,261) = 36.26	p=5.84e-9
3.1g	Visual threshold UST vs MST	Linear Mixed Model ANOVA (fixed effect: cohort)	12 UST sessions, 139 MST sessions	F(1,31) = 0.45	p=0.51
Not shown	Visual sensitivity UST vs MST	Linear Mixed Model ANOVA (fixed effect: cohort)	12 UST sessions, 139 MST sessions	F(1,31)=3.74	p=0.06
3.1h	Linear dependence between visual saliency and RT (UST)	Linear Mixed Model ANOVA (fixed effect: saliency, random effect: mouse ID)	493 trials	F(1,482) = 5.81	p=0.02
	Linear dependence between visual saliency and RT (MST)		3424 trials	F(1,3371) = 144.67	p=1.17e-32
	Linear dependence between auditory saliency and RT (MST)		4276 trials	F(1,4219) = 16.52	p=4.91e-05
3.1h	Auditory vs visual RT (MST)	Linear Mixed Model ANOVA (fixed effect: modality, controlling for saliency, random effect: mouse ID)	3424 visual trials, 4276 auditory trials	F(1,7599) = 706.89	p=5.28e-149
3.1h	Visual RT for UST vs MST	Linear Mixed Model ANOVA (fixed effect: cohort, controlling for saliency)	3917 trials (493 UST + 3424 MST) Subthreshold: 71 UST trials, 483 MST trials Threshold: 115 UST trials, 748 MST trials Suprathreshold: 130 UST trials, 933 MST trials Max: 168 UST trials, 1217 MST trials	F(1,3865) = 60.17	p=1.11e-14
3.2f	Maximal z-cored response early vs late phase per laminar zone	Linear Mixed Model ANOVA (fixed effect: temporal window, random effect: mouse ID)	Supragranular - 91 neurons	F(1,194) = 4.60	p=0.03
			Granular - 88 neurons	F(1,171) = 0.00	p=0.96
			Infragranular: 582 neurons	F(1,1284) = 23.32	p=1.53e-06
3.2f	Hit/miss modulation different across laminar zone	Linear Mixed Model ANOVA (fixed effect: laminar zone, random effect: mouse ID)	91+88+592=771 neurons	F(2,771) =4	p=0.019
				F(1,784)=12.97 F(1,784)=6.50 F(1,784)=0.58	G vs SG: p=0.0033 G vs IG: p=0.01 IG vs SG: p=0.45
3.4c	Fraction of neurons coding (significant AUC) different across layers	Linear Mixed Model ANOVA (fixed effect: laminar zone, random effect: mouse ID) Posthoc comparison (Linear hypothesis test on coefficients) (Only significant reported, rest p>0.05)	Visual orientation; no significant layer differences.		p>0.05
			Visual change occurrence; 4 sessions with enough supragranular neurons, 14 sessions with enough infragranular neurons,	F(1,16)=7.21	p=0.016
			Visual hit/miss threshold; 3 sessions with enough granular neurons, 15 sessions with enough infragranular neurons	F(1,15)=5.21	p=0.037
			Visual hit/miss maximal; 3 sessions with enough granular neurons, 15 sessions with enough infragranular neurons	F(1,15)=4.96	p=0.042
3.4d	Earliest time point of increase in coding fraction	Fraction significant neurons exceeds 2 std above baseline (-	NE: 116 neurons UST: 128 neurons	Z>2	p<0.02275

		1000 to 0 ms). This corresponds to a one-sided t-test with p < .02275. Very similar results were obtained with a threshold of 1 or 3 std above baseline.	MST: 306 neurons		
3.4d	For each variable, the difference in latency to coding between cohorts	Bootstrap test (n=1000 resamples). If the difference between bootstrap distribution exceeded the 97.5 percentile this was deemed significant. This corresponds to a two-sided p-value of 0.05.	NE: 116 neurons UST: 128 neurons MST: 306 neurons		Hit/miss coding latency threshold changes between UST and MST: P=0.012 Rest p>0.05.
	Subsample control (4d)	Bootstrap test on n=1000 resamples with the same number of neurons between UST and MST.	UST: 128 neurons MST: 128/306 neurons		Hit/miss coding latency threshold changes between UST and MST p<0.05. Rest p>0.05.
3.4e	Linear dependence earliest hit/miss coding of population activity and RT (on individual sessions)	Linear model ANOVA	26 sessions	F(1,24) = 5.15	p=0.03
3.4f	Linear dependence earliest increase in visual hit/miss coding and RT (on bootstrapped condition averages)	Linear model ANOVA	4 conditions (UST and MST, 2 saliencies each) RT boostrapped from n=1269 Vthr UST trials, n=1292 Vmax UST trials, n=960 Vthr MST trials, n=1051 Vmax MST trials Coding onset boostrapped from n= 128 UST V1 neurons, n=306 MST neurons.	F(1,2) = 102.33	p=0.0096
			Bootstrap results: (Mean and 95% CI)		Slope: 1.58 (0.27-2.52) Offset: -569 ms (-985 to -10).
3.5g,5h	Effect of inactivation on discrimination performance (d-prime) comparing early or late inactivation versus control trials for different saliencies, modalities, and cohorts	Linear Mixed Model ANOVA (fixed effect of inactivation, random effect: mouse ID) with Bonferroni-Holm correction for multiple comparisons (4 comparisons)	threshold visual change, UST, Ctrl vs Early 18 sessions	F(1,32)=16.71	p=0.0032
			threshold visual change, UST, Ctrl vs Late 18 sessions	F(1,32)=0.29	p=1
			threshold visual change, MST, Ctrl vs Early 34 sessions	F(1,59)=35.32	p=0.000002
			threshold visual change, MST, Ctrl vs Late 34 sessions	F(1,54)=13.90	p=0.00553
			maximum visual change, UST, Ctrl vs Early 18 sessions	F(1,32)=14.80	p=0.0064
			maximum visual change, UST, Ctrl vs Late 18 sessions	F(1,30)=1.19	p=0.85
			maximum visual change, MST, Ctrl vs Early 34 sessions	F(1,58)=32.56	p=0.000005
			maximum visual change, MST, Ctrl vs Late 34 sessions	F(1,53)=13.48	p=0.0067
			auditory change, MST, all comparisons 34 sessions	All F < 3	All p>0.1
3.5g,5h (not shown)	Subsample control of 5g,h	Same as above but for n=1000 resamples of 18/34 MST sessions to match # of UST session	threshold visual change, MST, Ctrl vs Late 18/34 MST sessions		P<0.05 for 82% of resamples
			maximal visual change, MST, Ctrl vs Late 18/34 MST sessions		P<0.05 for 86% of resamples
3.5i	Linear dependence median RT and percentage reduction d-prime	Linear Mixed Model ANOVA (fixed effect of reaction time, random effect: mouse ID)	Early silencing: 40 conditions (21 Thr and 19 Max)	F(1,33)=1.71	p=0.20

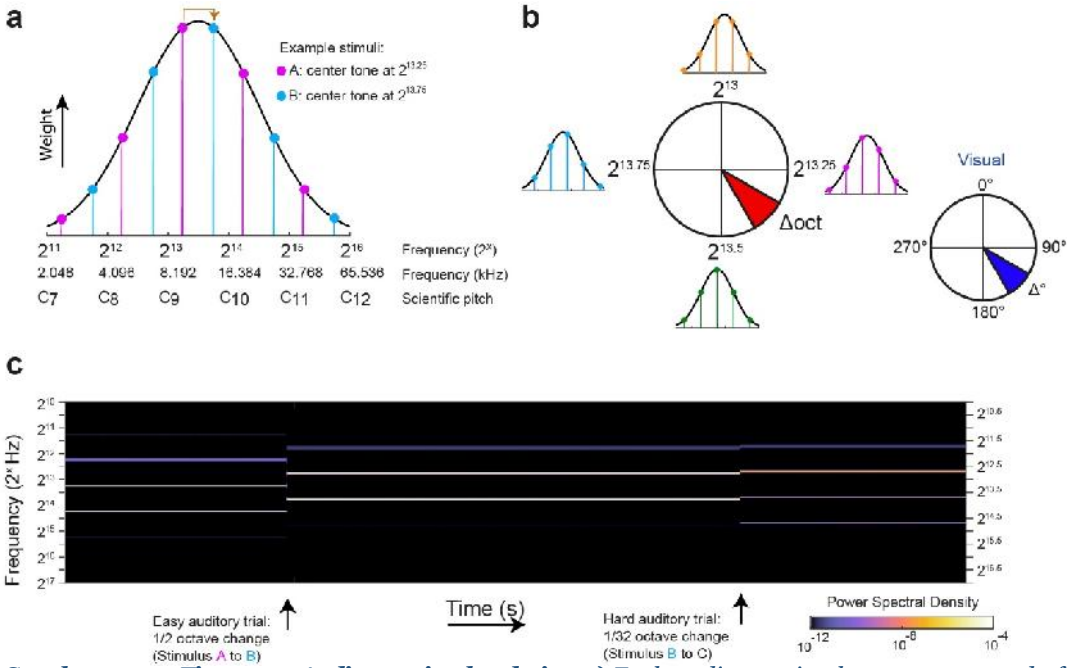
3.5j	Linear dependence median RT and percentage reduction d-prime	Linear Mixed Model ANOVA (fixed effect of reaction time, random effect: mouse ID)	Late silencing: 45 conditions (22 Thr and 23 Max)	F(1,15)=10.04	P=0.006
3.5j (not shown)	Linear dependence median RT and percentage reduction d-prime (controlling for visual dprime on control trials)	Linear Mixed Model ANOVA (fixed effect of reaction time and visual dprime; random effect: mouse ID)	Late silencing: 45 conditions (22 Thr and 23 Max)	RT: F(1,20)=11.77, Dprime: F(1,42)=1.93	p=0.003 p=0.172
3.6d	Effect of inactivation on D-prime. Comparing Early or Late inactivation versus control trials for different saliencies, modalities, sides, cohorts.	Linear Mixed Model ANOVA (fixed effect: inactivation, random effect: mouse ID) with Bonferroni-Holm correction for multiple comparisons (4 comparisons)	Visual contralateral threshold detection, UST, Ctrl vs Early 7 sessions	F(1,14)=24.57	P=6.3179e-04
			Visual contralateral threshold detection, UST, Ctrl vs Late 7 sessions	F(1,14)=2.15	P=0.1644
			Visual contralateral threshold detection, MST, Ctrl vs Early 6 sessions	F(1,12)=17.93	P= 0.0023
			Visual contralateral threshold detection, MST, Ctrl vs Late 7 sessions	F(1,14)=45.14	P=3.6441e-05
		Linear Mixed Model ANOVA (fixed effect: inactivation, random effect: mouse ID) with Bonferroni-Holm correction for multiple comparisons (4 comparisons)	Visual contralateral maximum detection, UST, Ctrl vs Early 7 sessions	F(1,14)=0.16	P=0.9656
			Visual contralateral maximum detection, UST, Ctrl vs Late 7 sessions	F(1, 9)=0.53	P=0.9656
			Visual contralateral maximum detection, MST, Ctrl vs Early 4 sessions	F(1, 8)=3.73	P=0.2636
			Visual contralateral maximum detection, MST, Ctrl vs Late 7 sessions	F(1,13)=6.38	P=0.1023
		Linear Mixed Model ANOVA (fixed effect: inactivation, random effect: mouse ID) with Bonferroni-Holm correction for multiple comparisons (4 comparisons)	Visual ipsilateral threshold detection, UST, Ctrl vs Early 7 sessions	F(1, 7)=0.04,	P=1
			Visual ipsilateral threshold detection, UST, Ctrl vs Late 7 sessions	F(1,16)=0.80	P=1
			Visual ipsilateral threshold detection, MST, Ctrl vs Early 2 sessions	F(1,10)=0.02	P=1
			Visual ipsilateral threshold detection, MST, Ctrl vs Late 7 sessions	F(1,10)=0.38	P=1
3.6e	Linear dependence between percentage reduction d-prime and reaction time	Linear Mixed Model ANOVA (fixed effect: median RT, random effect: mouse ID)	30 conditions (7 UST thr, 7 UST high, 9 MST low, 7 MST high)	F(1,26)=9.78	p=0.004 r^2= 0.7056
3.7b	Pre-stim (-500 to 0 ms) vs post-stim (0 to +500 ms) decoding improvement over chance	Linear Mixed Model ANOVA (fixed effect: time window, random effect: mouse ID)	11 sessions combined across cohorts	F(1,17)=44.76	p=4.118e-06
3.7c	Significant decrease in noise correlations versus baseline for visual trials split by choice	Linear Mixed Model ANOVA (fixed effect of choice; random effect: mouse ID)	NE – Miss trials, 2904 pairs	F(1,5805)=14.67	P<1e-4
			NE – Hit trials, 1476 pairs	F(1,2950)=0.33	p=0.56
			UST – Miss trials, 1930 pairs	F(1,3856)=0.02	p=0.88
			UST – Hit trials, 1930 pairs	F(1,3856)=82.44	p<1e-19
			MST – Miss trials, 13692 pairs	F(1,27467)=3.31	p=0.069

			MST – Hit trials, 13972 pairs	F(1,28188)=142.96	P<1e-33
3.7d	Difference in visual hit reaction time between cohorts (median RT of session)	Two-sided Wilcoxon rank-sum test	11 UST sessions vs 44 MST sessions		p=0.0041
3.7e	Earliest time point of decorrelation	Earliest time point that noise correlations drop below baseline minus two standard deviations values. This corresponds to a one-sided t-test with p<0.05. Similar results were obtained with more or less strict thresholds.	From 59 sessions: UST: 4730 neuron pairs MST: 17826 neuron pairs		Z<-2
3.7f	Linear dependence reaction time and earliest time point of decorrelation	Pearson correlation	6 condition averages (fast, mid and slow tertiles for UST and MST each)		r=0.960, p=0.002
3.7g	Difference in z-scored activity between hit-miss during 100-200 ms window	Linear Mixed Model ANOVA (fixed effect: hit/miss, random effect: mouse ID)	UST - Thr - 78 neurons	F(1,156)=10.16	P=0.002
			UST - Max - 78 neurons	F(1,152)=5.66	P=0.019
3.7h	Difference in z-scored activity between hit-miss during 100-200 ms window	Linear Mixed Model ANOVA (fixed effect: hit/miss, random effect: mouse ID)	MST - Thr - 134 neurons	F(1,268)=13.19	P=0.001
			MST - Max - 120 neurons	F(1,254)=0.59	P=0.445
3.7i	Difference in noise correlation for each time bin	Two-sided bootstrapped confidence interval test	230 UST and MST neurons, 1564 neuron pairs, 1000 bootstraps		Black lines in Figure 5i, indicate P<0.05
	Supplementary figures:				
S3.2f	Linear dependence between auditory d-prime and RT	Linear Mixed Model ANOVA (fixed effect: RT, controlling for saliency; random effect: mouse ID)	139 MST sessions, 4 saliencies each	F(1,298) = 10.43	p=0.00138
S3.2g	Linear dependence between visual d-prime and RT	Linear Mixed Model ANOVA (fixed effect: RT, controlling for saliency; random effect: mouse ID)	151 UST and MST sessions, 4 saliencies each	F(1,268) = 11.36	p=0.00086
S3.3b	Difference in trial-averaged z-scored firing rate in 0 to 200ms post-stimulus window between threshold and maximal visual change conditions per cohort	Linear Mixed Model ANOVA (fixed effect: saliency, random effect: mouse ID)	163 NE neurons	F(1,326) = 7.27	P=0.0079
			128 UST neurons	F(1,256) = 20.43	p=9.46e-06,
			525 MST neurons	F(1,1568) = 35.90	p=2.56e-09
S3.3i	Difference in trial-averaged z-scored firing rate in -300 to 300ms window between lick and no-lick conditions	Linear Mixed Model ANOVA: Fixed effect of Licking: F(1,1966)=379.35, p=2.13e-77; Fixed effect of Cohort: F(2,1966)=6.08, p=0.002; Interaction effect Licking * Cohort: (2,1966)=10.59, p=2.658e-05;	163 NE neurons, 128 UST neurons 525 MST neurons;	Posthoc comparison: Linear hypothesis test on coefficients	p-value in figure, *p<0.05, **p<0.01, ***p<0.001
			Specific posthoc comparison interaction term Licking*Cohort for NE versus MST&UST.	F(1,1960)=19.71,	p=9.526e-06
S3.3j	Difference in trial-averaged z-scored firing rate in -300 to 300ms window between hit and incorrect conditions for trained UST and MST conditions	Linear Mixed Model ANOVA (fixed effect: Correct, random effect: Mouse ID):	Visual Incorrect vs Hits, n = 128 UST neurons	F(1,256) = 2.59	p=0.109
			Visual Incorrect vs Hits, n = 525 MST neurons	F(1,1371) = 140	P=2.54e-08
			Auditory Incorrect vs Hits, n = 525 MST neurons	F(1,1371) = 9.67	p=0.002
S3.4a	Difference in explained variance over single trials between cohorts over all trials using all predictors	Linear Mixed Model ANOVA (Fixed effect: cohort): (F(2,516) = 4.71, p=0.01, ANOVA) Posthoc comparison: Linear hypothesis test on coefficients	n=116 NE neurons, n=128 UST neurons, n=272 MST neurons, (only neurons >0.5 Hz) NE vs UST	F(1,513)=0.81	p=0.36
			NE vs MST	F(1,513)=8.28	p=0.004
			UST vs MST	F(1,513)=3.62	p=0.058
S3.4b	Difference in explained variance over averaged trial types between cohorts over all trials using all predictors	Linear Mixed Model ANOVA (fixed effect: cohort): (F(2,516) = 7.01, p=0.001, ANOVA) Posthoc comparison: Linear hypothesis test on coefficients	(n same as Fig. S4a) NE vs UST	F(1,513)=1.40	p=0.23
			NE vs MST	F(1,513)=12.55	p=0.0004
			UST vs MST	F(1,513)=5.06	p=0.025
S3.4c	Explained variance averaged over 0-200ms	Linear Mixed Model ANOVA (fixed effect: cohort with posthoc	n=116 NE neurons n=128 UST neurons	Vision : (MST vs	Vision : (MST vs NE)

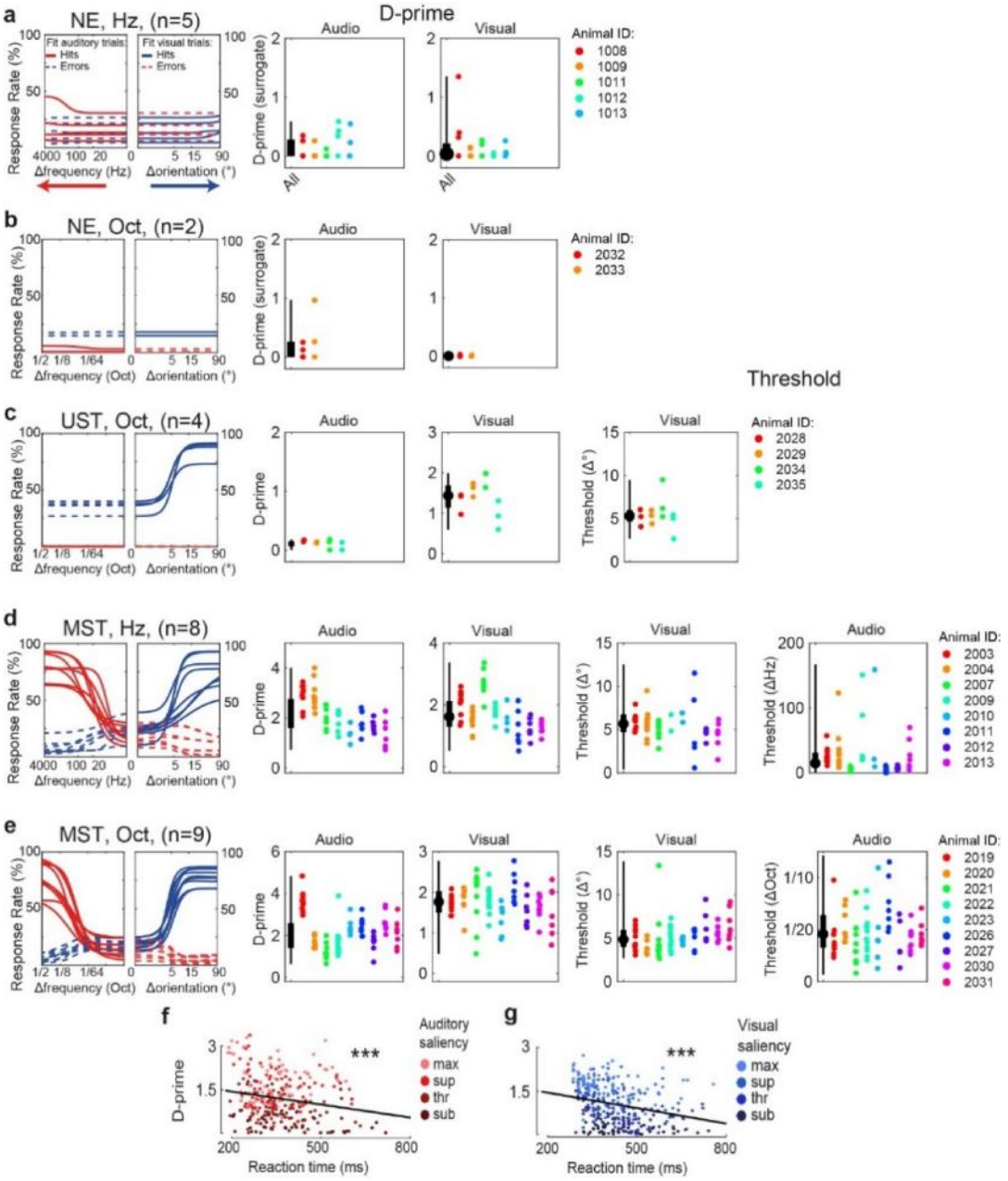
	(early activity) between predictor sets within cohort	comparison: Linear hypothesis test on coefficients)	n=272 MST neurons	NE):F(1,513)=16.60 (NE vs UST):F(1,513)=22.40 (MST vs UST):F(1,513)=2.09 Movement: (MST vs NE):F(1,513)=23.16, (NE vs UST):F(1,513)=6.28, (MST vs UST):F(1,513)=3.93, Hit: (UST vs NE):F(1,513)=39.48 (MST vs NE):F(1,513)=2.02 (MST vs UST):F(1,513)=23.06 Pupil: (MST vs NE):F(1,513)=0.17 (NE vs UST):F(1,513)=3.05 (MST vs UST):F(1,513)=6.32	p=0.000053 (NE vs UST) p=0.0000028 (MST vs UST) p=0.14888 Movement: (MST vs NE) p=0.0000019 (NE vs UST) p=0.0125 (MST vs UST) p=0.0480 Hit: (UST vs NE) p=0.0000000007 (MST vs NE) p=0.1562 (MST vs UST) p=0.000002 Pupil: (MST vs NE) p=0.6800447193 (NE vs UST) p=0.0815248918 (MST vs UST) p=0.0122326371
S3.4d	Explained variance averaged over 200-1000ms (late activity) between predictor sets within cohort	Linear Mixed Model ANOVA (fixed effect: cohort with posthoc comparison: Linear hypothesis test on coefficients)	n=116 NE neurons n=128 UST neurons n=272 MST neurons	Vision : (MST vs NE):F(1,513)=9.14 (NE vs UST):F(1,513)=12.27 (MST vs UST):F(1,513)=1.13 Movement: (MST vs NE):F(1,513)=32.08 (NE vs UST):F(1,513)=7.45 (MST vs UST):F(1,513)=6.74 Hit: (MST vs NE):F(1,513)=53.50 (NE vs UST):F(1,513)=41.98 (MST vs UST):F(1,513)=0.03 Pupil: (MST vs NE):F(1,513)=0.15 (NE vs UST):F(1,513)=1.27 (MST vs UST):F(1,513)=0.91	Vision : (MST vs NE) p=2.632963e-03 (NE vs UST) p=5.003239e-04 (MST vs UST) p=2.885020e-01 Movement: (MST vs NE) p=2.470784e-08 (NE vs UST) p=6.573378e-03 (MST vs UST) p=9.705983e-03 Hit: (MST vs NE) p=1.001216e-12 (NE vs UST): p=2.162020e-10 (MST vs UST): p=8.557330e-01 Pupil: (MST vs NE): p=7.017013e-01 (NE vs UST) p=2.596394e-01 (MST vs UST) p=3.409608e-01
S3.4e	Onset latency of significant hit encoding different for UST and MST cohorts Onset latency = time bin where EV exceeded baseline + 2 standard deviations	Linear Mixed Model ANOVA (fixed effect: modality)	n=128 UST neurons n=272 MST neurons	F(1,337) = 1.54	p=0.21
S3.6b	Effect of inactivation on discrimination performance (d-prime) comparing early or late inactivation versus control trials for auditory saliencies	Linear Mixed Model ANOVA (fixed effect of inactivation, random effect: mouse ID)	Auditory change, MST, all comparisons 34 sessions	All F < 6	All p>0.1
S3.6e	Effect of inactivation on discrimination performance (d-prime) comparing early or late inactivation versus control trials for visual saliencies, for sessions with the fastest average RT (top 50%)	Linear Mixed Model ANOVA (fixed effect of inactivation, random effect: mouse ID)	n=18 sessions	Visual thr change MST, Ctrl vs Early: F(1,31)=29.15	p=0.000027
				Visual max change MST, Ctrl vs Early:: F(1,25)=46.00	p=0.000002
				Visual thr change MST, Ctrl vs Late: F(1,27)=2.18	p=0.606787
				Visual max change MST, Ctrl vs Late: F(1,28)=9.05	p=0.022190

S3.6h	Effect of inactivation on discrimination performance (d-prime) comparing early or late inactivation versus control trials for visual saliencies, for sessions with the slowest average RT (bottom 50%)	Linear Mixed Model ANOVA (fixed effect of inactivation, random effect: mouse ID)	n=17 sessions	Visual thr change MST, Ctrl vs Early: F(1,30)=11.76	p=0.007136
				Visual max change MST, Ctrl vs Early: F(1,30)=5.34	p=0.111827
				Visual thr change MST, Ctrl vs Late: F(1,31)=16.31	p=0.001313
				Visual max change MST, Ctrl vs Late: F(1,29)=6.98	p=0.0047151
S3.6k	Linear dependence median RT and percentage reduction in criterion	Linear Mixed Model ANOVA (fixed effect of RT, random effect: mouse ID)	87 conditions (45 Thr and 42 Max)	F(1,39)=1.55	p=0.22
S3.6fk (not shown)	Linear dependence median RT and FA visual lick spout	Linear Mixed Model ANOVA (fixed effect of RT, random effect: mouse ID)	94 conditions (47 Thr and 47 Max)	F(1,46)=0.05	p=0.82
S3.6l	Late silencing delays reaction times: difference in reaction time between control and late silencing visual hits	Generalized Linear Mixed Model ANOVA (fixed effect of inactivation, random effect: mouse ID)	maximal changes: Control hits (n=795 trials) and late silencing hits (n=388 trials)	F(1,1181)=3.12	p=0.08
			threshold changes: Control hits (n=509 trials) and late silencing hits (n=252 trials)	F(1,759)=0.28	p=0.59
S3.7b-c	D-prime visual and audio change, UST and MST, Ctrl vs Early and Ctrl vs Late (S1 inactivation)	Linear Mixed Model ANOVA (fixed effect of inactivation, random effect: mouse ID) with Bonferroni-Holm correction for multiple comparisons (4 comparisons)	13 UST sessions, 16 MST sessions	All F < 6	Non-significant, all comparisons p>0.05
S3.8a	D- prime maximum, visual UST vs MST	Linear Mixed Model ANOVA (fixed effect: Cohort, controlling for side).	UST: 4 mice x 7 sessions x 2 sides MST: 4 mice x 15 sessions x 2 sides	F(1,44)=3.1822	P=0.0814
S3.8b	Visual detection threshold, UST vs MST	Linear Mixed Model ANOVA (fixed effect: Cohort, controlling for side)	MST: 4 mice x 2 sides UST: 4 mice x 2 sides	F(1,16)=0.3675	p=0.5529
S3.8c	Visual RT, UST vs MST	Generalized Linear Mixed Model ANOVA (fixed effect: cohort, controlling for saliency and side) with Bonferroni-Holm correction for multiple comparisons (2 comparisons)	1395 trials	F(1,1391)= 54.075	p=3.2832e-13
	Visual RT (UST), Thr vs Max	Generalized Linear Mixed Model ANOVA (fixed effects: Saliency, controlling for Side, random effects: session, mouse) with Bonferroni-Holm correction for multiple comparisons (2 comparisons)	617 trials	F(1,614)= 129.85	P=4.0354e-27
	Visual RT (MST), Thr vs Max		778 trials	F(1,775)= 56.919	P=2.5486e-13
	Tactile RT (MST), Thr vs Max	Generalized Linear Mixed Model ANOVA (fixed effects: Saliency, controlling for Side, random effects: session, mouse)	625 trials	F(1,622) = 5.403	P= 0.020424
S3.8d	D-prime tactile contralateral, threshold, MST, Ctrl vs Early	Linear Mixed Model ANOVA (fixed effect: inactivation, random effect: mouse ID) with Bonferroni-Holm correction for multiple comparisons (2	6 sessions	F(1,12)=3.78	p=0.1513
	D-prime tactile contralateral, threshold,		9 sessions	F(1,14)=0.12	p=0.7339

	MST, Ctrl vs Late	comparisons)			
S3.8e	Effect of inactivation on the percentage of right-sided licks for condition: Thr Contralateral Visual	Linear Mixed Model ANOVA (fixed effect: inactivation, random effect: mouse ID) with Bonferroni-Holm correction for multiple comparisons (4 comparisons)	UST, Early vs Ctrl, 7 sessions	F(1,10.7)=39.08	P = 2.0943e-04
			UST, Late vs Ctrl, 7 sessions	F(1,10.7)=7.805	P = 0.0178
			MST, Early vs Ctrl, 7 sessions	F(1,12)=32.456	P = 2.0943e-04
			MST, Late vs Ctrl, 6 sessions	F(1,18)=38.242	P = 3.0993e-05
S3.9a	Significant decrease in noise correlations versus baseline for audio trials split by choice	Linear Mixed Model ANOVA (fixed effect of choice, random effect: mouse ID)	NE – Miss trials, 2904 pairs	F(1,5805)=1.96	P=0.16
			NE – Hit trials, 2106 pairs	F(1,1748)=3.83	p=0.054
			MST – Miss trials, 13656 pairs	F(1,27395)=22.61	p=1.99e-6
			MST – Hit trials, 14462 pairs	F(1,28847)=99.90	P=1.7e-23
S3.9b	Linear dependence reaction time and earliest time point of decorrelation relative to first lick	Pearson correlation	6 condition averages (fast, mid and slow tertiles for UST and MST each)	r=0.738	p=0.094

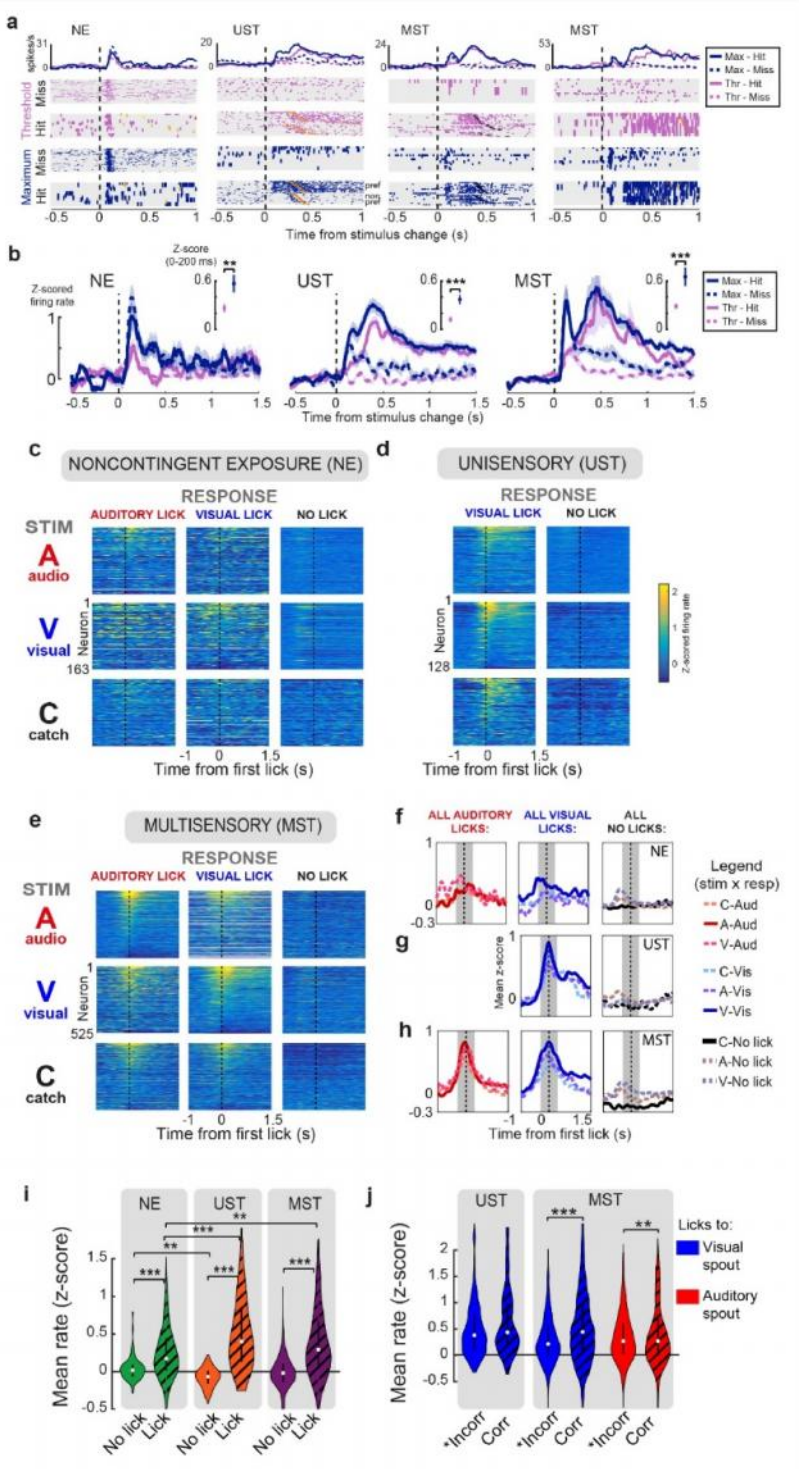


Supplementary Figure 3.1 Auditory stimulus design a) *Each auditory stimulus was composed of five pure tones at harmonic frequencies (octaves below and above other tones). The weight with which each tone contributed to the overall stimulus was taken from a Gaussian distribution across all possible tones. The example stimulus A in pink is composed of a tone of $2^{13.25}$ Hz (center tone, highest weight) and two lower (at $2^{11.25}$ and $2^{12.25}$ Hz) and two higher harmonics (at $2^{14.25}$ and $2^{15.25}$ Hz). Tones followed scientific pitch and are expressed as powers of two: 2^{13} corresponds to 8.192 kHz, and C_9 in scientific pitch notation. During an auditory trial, the stimulus changed to a stimulus of five new harmonic tones with different weights (for example stimulus A to B).* ***b)*** *The center diagram shows the circular arrangement of all stimuli. For each cardinal direction the insets show the tonal weights associated with these stimuli. Note how ever increasing the center tone frequency ultimately results in a circular shift back to the starting stimulus. This feature is exploited in the Shepard illusion, but note that our stimuli were static and had no illusory component. This circularity can also be seen in panel a: going up and down half an octave from stimulus A always results in stimulus B. The auditory stimulus set is therefore circular and mirrors the visual stimulus set with drifting gratings in all orientations (inset for comparison in right lower corner). The amount of frequency change (expressed in partial octaves, red) or orientation change (expressed in degrees, blue) determined saliency.* ***c)*** *Spectrogram over time including two auditory change trials. Auditory stimuli continued to be presented until the next auditory change, which could be identified based on a difference in spectral content, and experienced as an increase or decrease in pitch. The example shows an easy auditory trial (salient change; stimulus A to B, half an octave) followed later by a difficult trial (subtle change; 1/32 of an octave). Note that this is only a schematic depiction, hence time is depicted in arbitrary units.*

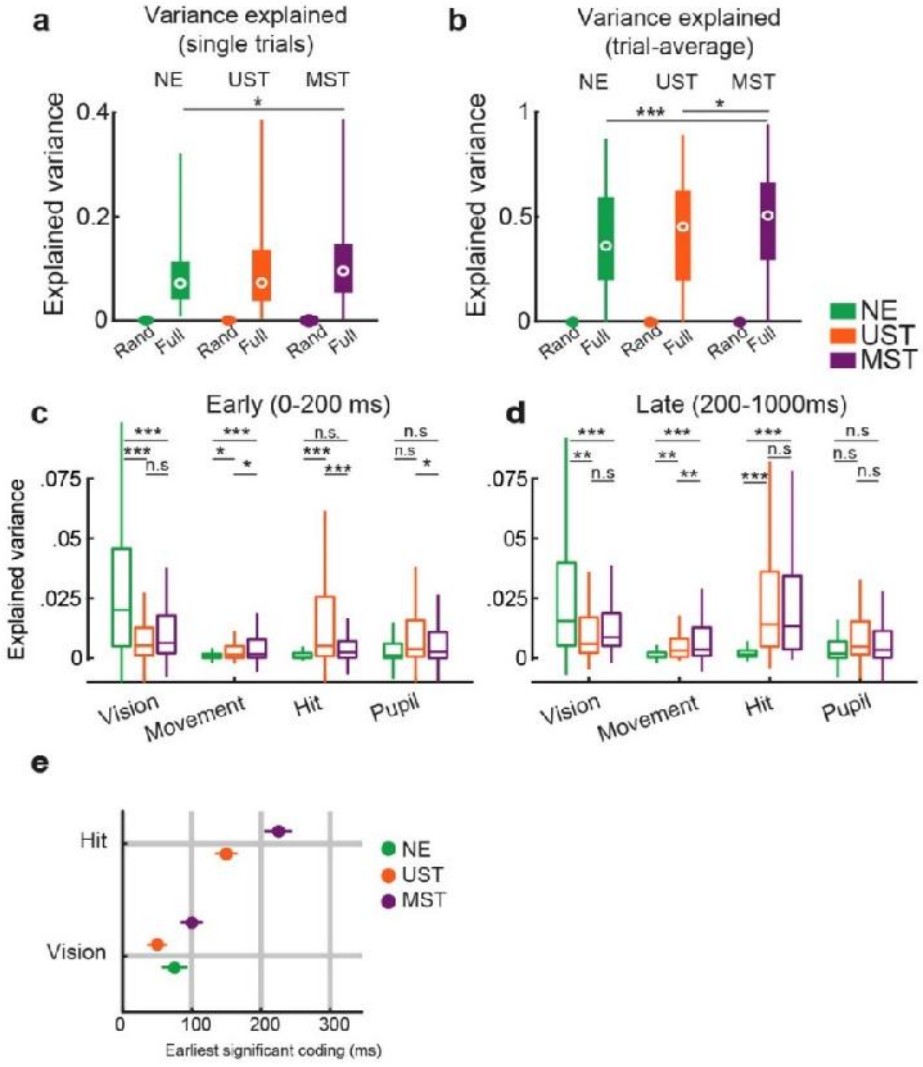


Supplementary Figure 3.2 Detailed psychophysical performance across versions of task A. *This figure shows the data for individual animals and individual sessions for each variant of task A. We implemented two versions of auditory stimuli (frequency changes expressed in the amount of Hertz, or octaves, see Methods) and split figures here based on version. The figure follows the same conventions as Fig. 1c-e in the main text. Dots in D-prime graphs indicate individual sessions from a single mouse (identified by color).* ***a)*** *Animals (N=5) were used for the noncontingently exposed (NE) variant with auditory changes in Hz. The two leftmost panels show fitted psychometric*

functions (two-alternative detection model) displaying behavioral response rates at increasing levels of auditory change (left panel) and increasing levels of visual orientation change (right panel). Solid lines are hits and dotted lines are errors. Blue indicates responses to the visual lick spout and red responses to the auditory lick spout. These baseline lick rates after stimulus changes in NE mice result from the animal spontaneously licking (some licks were rewarded, but this was not temporally related to the stimuli, see Methods). Sometimes licks were emitted accidentally, briefly after a change in stimulus ('surrogate hits'). The subpanels to the right side show for each animal (different rainbow colors) and for each session (single dots) the parameters of the single session fits for the asymptotic d-prime for auditory detection (left) and visual detection (right). Within each subpanel, the boxplot shows the median (dot), interquartile range (box limits) and minima and maxima (whiskers). ***b-e)*** *Same as (**a**), but for the other reward contingencies for task A (UST and MST). For animals trained to report either visual or auditory changes, the detection thresholds are also shown. These include the visual threshold in UST and the visual and auditory threshold in MST mice. Thresholds for non-rewarded conditions were higher than the maximum saliency or infinite. No mice were trained for the UST variant of the task with auditory changes in Hz.* ***f)*** *In MST animals, d-prime decreases as a function of reaction time for auditory conditions ($F(1,298) = 10.43$, $p=0.00138$; ANOVA). Each dot is one saliency condition within a single session.* ***g)*** *Same as (**f**), but for the negative correlation between reaction time and d-prime in all visual conditions across combined UST and MST sessions ($F(1,268) = 11.36$, $p=0.00086$; ANOVA).*

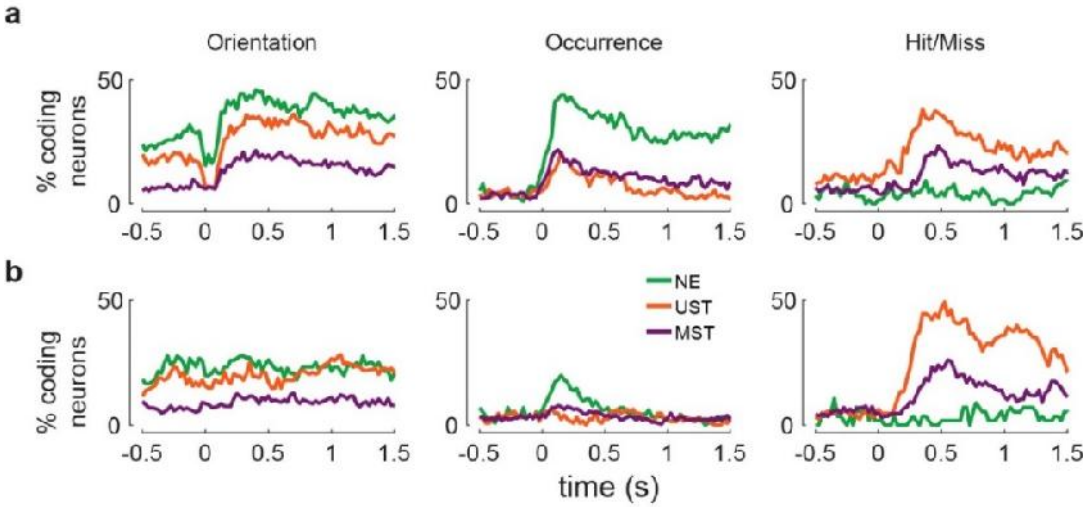

a
NE
UST
MST
MST
spikes/s
Threshold
Miss
Hit
Maximum
Miss
Hit
Max - Hit
Max - Miss
Thr - Hit
Thr - Miss
Time from stimulus change (s)
b
Z-score (0-200 ms)
Z-scored firing rate
NE
UST
MST
Time from stimulus change (s)
c
NONCONTINGENT EXPOSURE (NE)
RESPONSE
AUDITORY LICK
VISUAL LICK
NO LICK
STIM
A audio
V visual
C catch
Neuron
Time from first lick (s)
d
UNISENSORY (UST)
RESPONSE
VISUAL LICK
NO LICK
Neuron
Z-scored firing rate
Time from first lick (s)
e
MULTISENSORY (MST)
RESPONSE
AUDITORY LICK
VISUAL LICK
NO LICK
STIM
A audio
V visual
C catch
Neuron
Time from first lick (s)
f
ALL AUDITORY LICKS:
ALL VISUAL LICKS:
ALL NO LICKS:
NE
g
UST
Mean z-score
h
MST
Time from first lick (s)
Legend (stim x resp)
C-Aud
A-Aud
V-Aud
C-Vis
A-Vis
V-Vis
C-No lick
A-No lick
V-No lick
i
NE
UST
MST
Mean rate (z-score)
No lick
Lick
j
UST
MST
Mean rate (z-score)
Licks to:
Visual spout
Auditory spout
Incorr
Corr

Supplementary Figure 3.3 Early and late wave dynamics across levels of change saliency and aligned to lick (Task A). ***a*** *Raster plots for four example neurons showing early sensory-driven transients and late activity both for threshold and maximal orientation changes. Orange ticks indicate first lick and immediate reward delivery. Same conventions as in Fig. 2.* ***b*** *Averaging Z-scored firing rate across all neurons again reveals early sensory-driven activity in all cohorts, while hits are associated with a strong late increase in activity only in UST and MST mice for both visual saliency levels. The amplitude of early sensory-driven activity (average activity 0-200 ms, hits and misses combined) was smaller for threshold than for maximal changes for all three cohorts (shown in insets; NE: $F(1,326)=7.27$, $p=0.00739$; UST: $F(1,256)=20.43$, $p=9.47x10^{-6}$; MST: $F(1,1568)=35.90$, $p=2.57x10^{-9}$; Linear Mixed Model ANOVA). Lines and shading indicate mean ± SEM across neurons.* ***c-e*** *We computed the average z-scored activity across all recorded neurons in V1 aligned to the first lick for 9 stimulus-response combinations: three stimulus type conditions (A=auditory, V=visual, and C=catch – i.e. no change) and three response options (lick to visual spout, lick to auditory spout, and no lick). In NE mice, all conditions with licking (left 6 heatmaps) showed slight modulations of activity around licks, which were absent in the three conditions without licking (right 3 heatmaps). This lick-aligned modulation, however, was much more prominent in MST and UST mice (panels* ***(d)*** *and* ***(e)****, respectively). For trials in which there was no lick, activity was aligned to the median response latency from all other trials. Conditions for which not enough trials were present to compute a reliable mean z-score for that neuron (fewer than 3 trials), were omitted from the heatmap. Therefore, trial-averaged estimates for licks to the auditory spout are absent in UST animals (trained on vision only), as animals rarely responded to the never-rewarded auditory lick spout.* ***f-h*** *All conditions with licking (left 2 panels for each row) show activity modulations hundreds of milliseconds before and after lick-onset. Each plot combines three conditions from the heatmaps in c-e (taken per column), and shows the Z-scored firing rate averaged over neurons (f: NE; g: UST; h: MST). Lines show mean across neurons. As in* ***(d)****, not enough events were available to calculate licks to the auditory lick spout in UST animals.* ***i*** *Licks evoked consistently higher V1 firing activity than no-licks for all cohorts (ANOVA, n=163+128+525 neurons, $F(1,1966)=379$, $p=2.13x10^{-7}$) in the time window around lick onset (-300 ms to +300 ms relative to lick onset, gray-shaded patch in* ***(f-h)****). Lick modulation was stronger for trained cohorts versus naive mice (ANOVA, UST vs NE, MST vs NE, $F(2,1966)=6.08$, $p=0.002$); *$p<0.05$, **$p<0.01$, ***$p<0.001$. Violinplots show distribution of rates across neurons. Inner white dot and black line show median and interquartile range.* ***j*** *Correct licks were associated with higher V1 firing activity than incorrect licks in MST mice (ANOVA; UST, n=128, $F(1,256)=2.59$, $p=0.109$; MST n=525, visual, $F(1,1371)=140$, $p=2.54x10^{-8}$, auditory, $F(1,1371)=9.67$, $p=0.002$). *Incorrect licks include both false alarms and mistaken licks to opposite spout (e.g. visual stimulus, lick to auditory lick spout). Conditions are separated to allow comparing between licks to the same spout. Same style as* ***(i)****.*

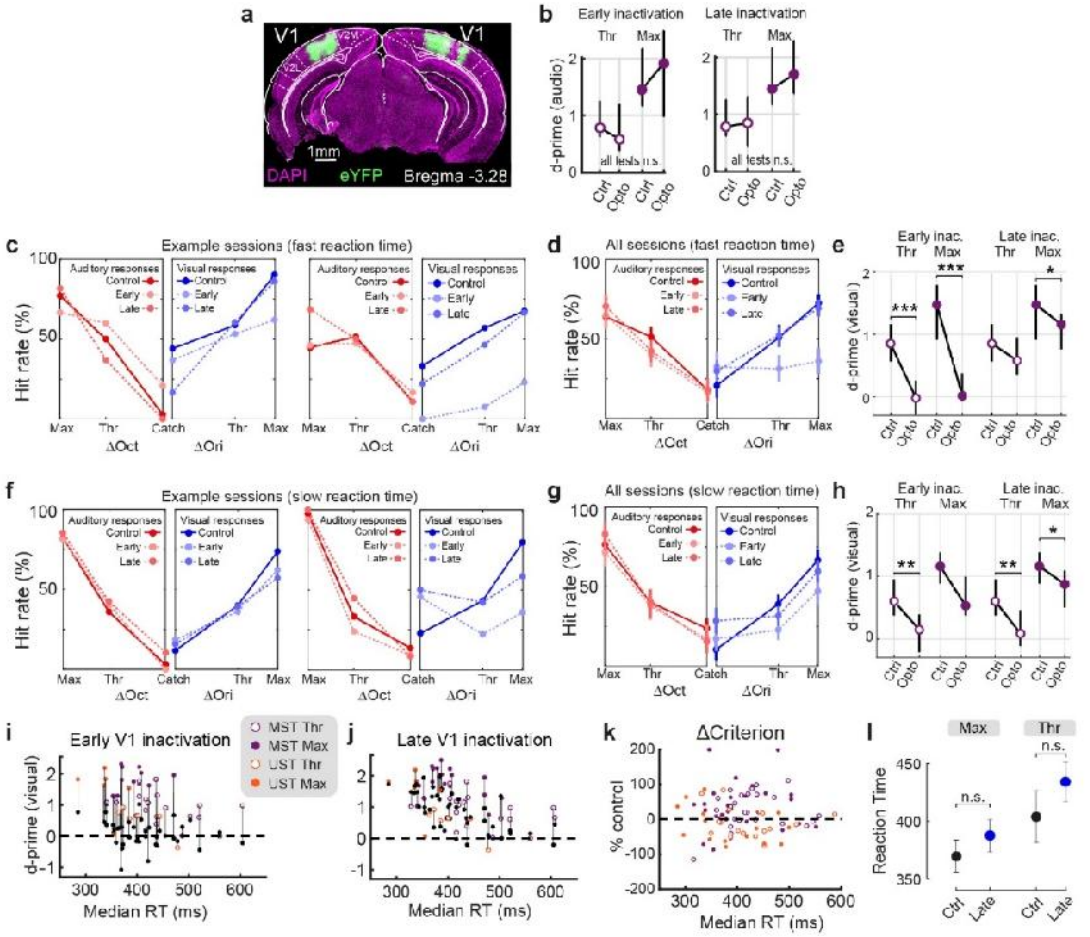


Supplementary Figure 3.4 Application of the generalized linear encoding model to explain variance of firing rates from different cohorts (Task A). a *We constructed two models. The first model operated as a null model and consisted of a random predictor only (Rand). The second model included all predictors (Full). We quantified the model performance in two ways. First, we computed the cross-validated explained variance (EV) over all single-trial firing rates for all neurons (n=116 NE neurons, n=128 UST neurons, n=272 MST neurons). The full model always explained significantly more variability than the random model (all p<1e-20), and explained variance was slightly higher for V1 neurons from MST mice than NE neurons (Linear Mixed Model ANOVA; F(2,516) = 4.71, p=0.01, ANOVA; Posthoc comparisons: MST vs NE: F(1,513)=8.28, p=0.004. NE vs UST: F(1,513)=0.81, p=0.36; MST vs UST: F(1,513)=3.62, p=0.058. Boxplots show the median (dot), interquartile range (box limits) and minima and maxima (whiskers).* ***b*** *We also quantified model performance by computing the EV of the firing rate averaged over the five stimulus x choice conditions that captured nearly all trials[20,38]. Again, the full model explained*

significantly more variability than the random model (all p<1e-33), and explained variance was higher for MST compared to NE and UST neurons (Linear Mixed Model ANOVA: F(2,516) = 7.01, p=0.001, ANOVA; Posthoc comparison: MST vs NE: F(1,513)= 12.55, p<0.001; NE vs UST: F(1,513)=1.40, p=0.23; MST vs UST: F(1,513)=5.06, p=0.025. Same style as a. ***c*** *We computed the variance of the firing rate as explained by each subset of predictors for each of the task versions over time (cf. Fig. 3c). During the early post-stimulus window (averaging EV over 0-200 ms) visual predictors explained more variance in NE than UST and MST mice. (ANOVA, n=116 NE neurons, 128 UST neurons, 272 MST neurons, *p<0.05, **p<0.01, ***p<0.001). Boxplots show the median and interquartile range (box limits) and 1 x interquartile range (whiskers).* ***d*** *Same as (**c**), but for the late window (200-1000 ms). The hit predictor now explains more variance than during the early period and explains more variance in both UST and MST than in NE mice (p<0.001). Therefore, knowing when licks occurred and whether the trial was a hit or not, contributed to predicting late V1 firing in UST and MST mice. Moreover, visual predictors continue to make strong contributions in the late phase across the three training cohorts.* ***e*** *The onset latency for significant EV by visual and hit predictors. Note the difference in onset latency of hit coding between UST and MST (corresponding to the ROC analysis in Fig. 4d). Shown are mean ± SEM across neurons. (UST n=128, MST n=272 neurons, ANOVA, F(1,337) = 1.54, p=0.21).*

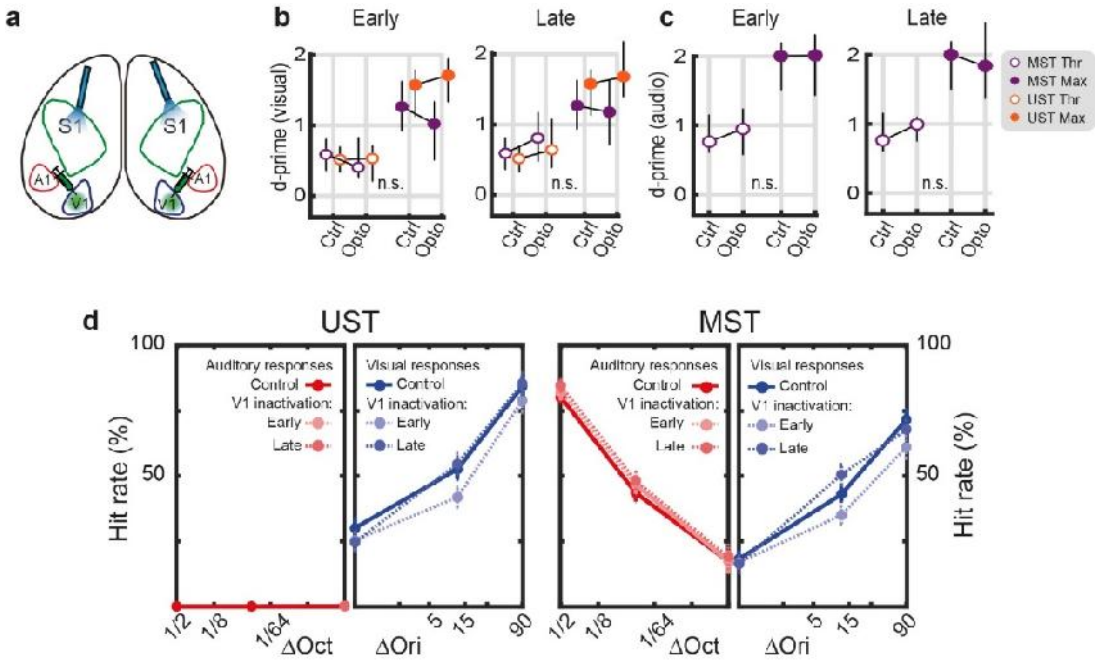


Supplementary Figure 3.5 Single neuron coding over time (Task A). ***a*** *Percentage of V1 neurons significantly encoding selected variables over time for visual changes of maximal saliency. The strength of encoding (AUC value above shuffled) gave similar results as the fraction of coding neurons (shown here).* ***b*** *Same as (**a**), but for threshold visual changes.*

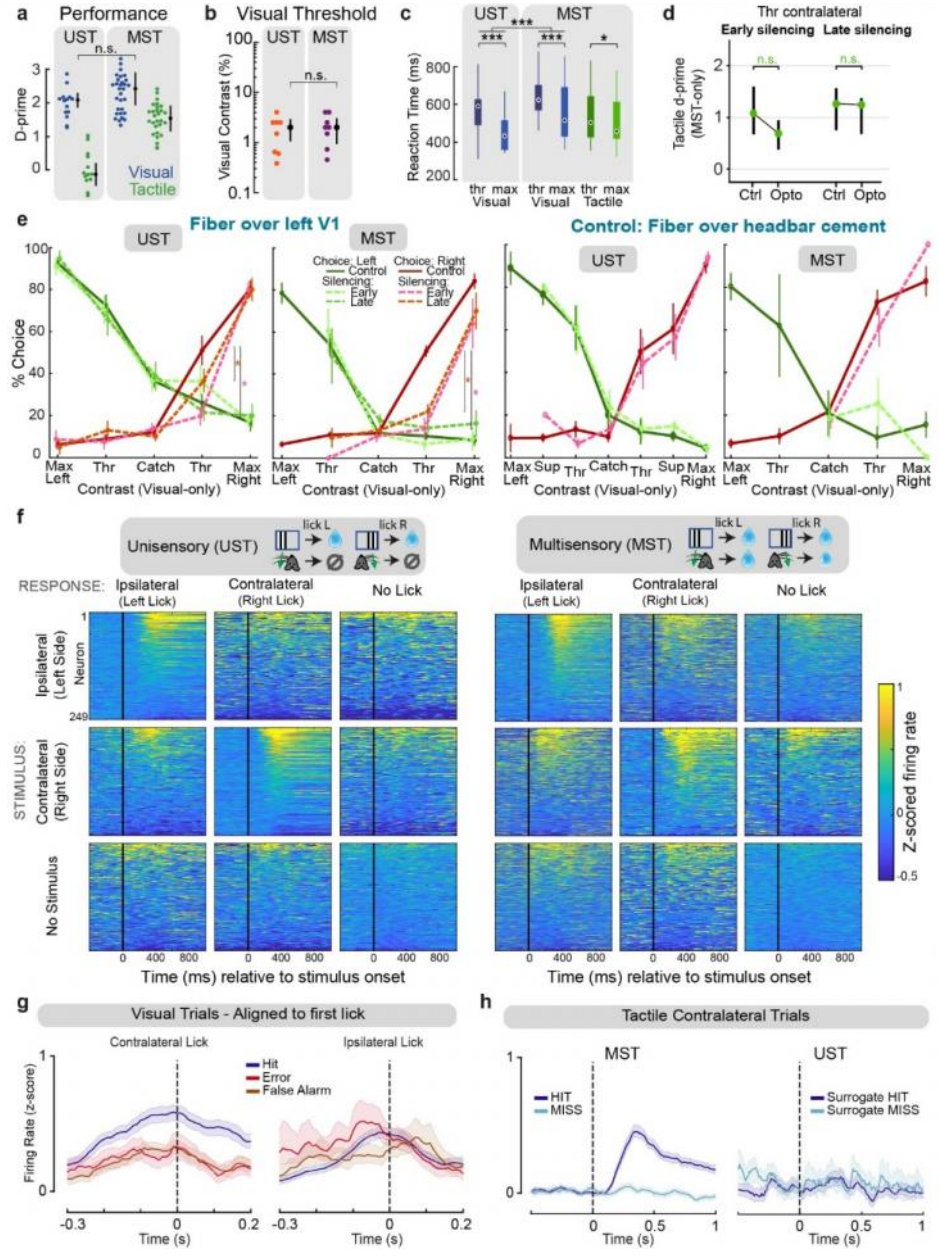


Supplementary Figure 3.6 Detailed characterization of early and late V1 inactivation in task A. ***a*** *Coronal histological section revealing localized bilateral expression in V1. V2L: lateral secondary visual cortex. V2M: medial secondary visual cortex. Histological analyses were repeated with comparable results for all 28 mice.* ***b*** *Neither early nor late silencing affected auditory discrimination performance (d-prime) for both threshold and maximum saliencies and across UST and MST cohorts (n=34 sessions, ANOVA, all F<6, all p>0.1).* ***c*** *Behavioral detection rates for two example sessions from MST mice with fast reaction times (median visual hit reaction time 359 and 362 ms).* ***d*** *Same as Fig. 5f, but for sessions with fast reaction times (top half of all reaction times).* ***e*** *Effect of early and late inactivation on d-prime for fast sessions, as a function of visual salience (thr vs. control and max vs. control. Asterisks indicate the result of a Linear Mixed Model ANOVA: *p<0.05, **p<0.01,***p<0.001. Exact p values: Visual thr change, Ctrl vs. Early: p=0.000027; Visual max change, Ctrl vs. Early: p=0.000002; Visual thr change, Ctrl vs. Late: p=0.606787; Visual max change, Ctrl vs. Late: p=0.022190.* ***f*** *Same as (**d**) for two example sessions from MST mice with slow reaction times (median visual hit reaction time 435 and 463 ms).* ***g*** *Same as (**d**) but for sessions with slow reaction times (bottom half of all reaction times).* ***h*** *Same as (**e**) but for sessions with slow reaction times. Exact p values: Visual thr change, Ctrl vs. Early: p=0.007136; Visual max change, Ctrl vs. Early: p=0.111827; Visual thr change, Ctrl vs. Late: p=0.001313; Visual max change, Ctrl vs. Late: p=0.0047151. Panels (**c-h**) jointly show how late silencing affects behavioral performance in slow sessions, but not fast sessions.* ***i*** *Scatter plot of visual d-prime on*

control (colored) and photostimulation trials (black). One data point is one session. Data points from the same session are connected with a line to visualize the reduction in d-prime. ***j*** *Same as (****i****) but for late silencing. Note how sessions with short reaction time are proportionally less affected than sessions with long reaction time. Quantification of this effect as the percentage reduction in d-prime is in Fig. 3i-j.* ***k*** *We tested whether late inactivation could affect motivation by changing the criterion parameter in our signal detection framework (see Methods). The reduction in visual criterion by late photostimulation was not significantly correlated to the median reaction time on control trials in the same recording session ($F(1,39)=1.55$, $p=0.22$). Similarly, we found no effect when we repeated this analysis on the false alarm rate directly ($F(1,46)=0.05$, $p=0.82$).* ***l*** *As late photostimulation partially reduced hit rate for visual changes in MST mice depending on reaction time (Fig. 3h, j), some visual changes were still detected. Visual hits with and without late photostimulation were not associated with a significant difference in reaction times (MST – max: $n=1185$ trials, $F(1,1179)=2.95$, $p=0.09$; MST – thr: $n=761$ trials, $F(1,761)=0.28$, $p=0.60$; Linear Mixed Model ANOVA). Traces show mean ± SEM across visual hits. *$p < 0.05$.*

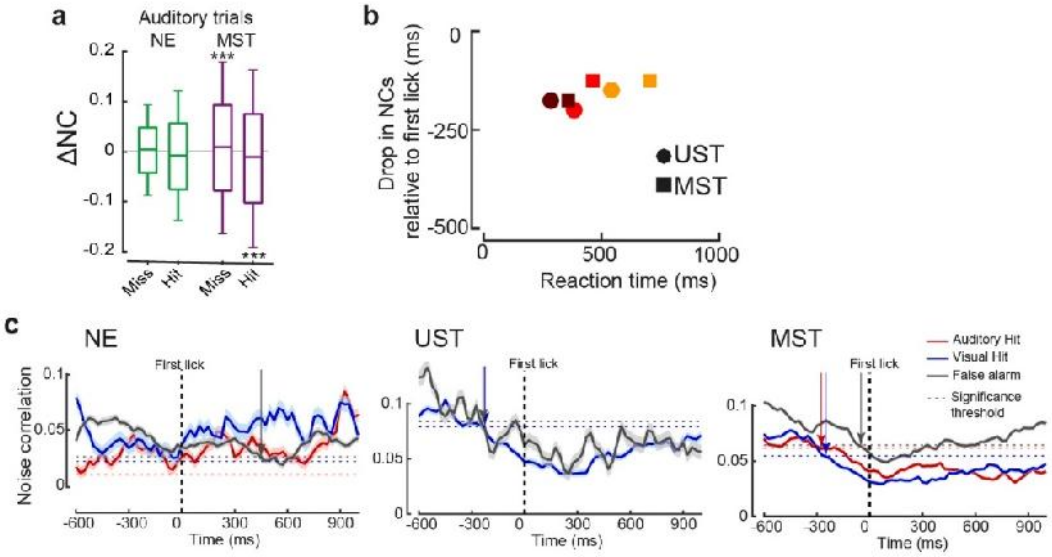


Supplementary Figure 3.7 Illumination of control area S1 in task A has no behavioral effects. a *Control experiment with positioning of the optical fiber over uninfected S1. **b** D-prime across visual conditions. Neither early nor late S1-illumination significantly affected visual detection performance for UST or MST mice (ANOVA, n= 29 sessions, all F<6, all p>0.05, corrected for 4 multiple comparisons (Bonferroni-Holm)). **c** Same as (**b**), but for auditory conditions. Neither early nor late S1-illumination significantly affected auditory detection performance (n= 29 sessions, all p>0.05). For **b** and **c**, errorbars denote inter-quartile range. **d** Behavioral response rates for UST (left) and MST (right) mice for control, early and late S1-illumination trials.*



Supplementary Figure 3.8 Effects of optogenetic V1 silencing on visuotactile behavior (task B). ***a*** *D-prime at maximum saliency for visual and tactile detection. Each dot represents one session and either right or left side detection performance. Visual performance was comparable for UST and MST mice (ANOVA, n=22 sessions, 8 mice, F(1,44)=3.18, p=0.0814). Note that null d-prime for tactile detection is expected for UST mice. For panels* ***a****,* ***b****,* ***d****,* ***e****, errorbars denote inter-quartile range.* ***b*** *Visual contrast detection thresholds were comparable for UST and MST mice (ANOVA, n=8 mice, F(1,16)=0.3675, p=0.5529). Computed for each mouse from psychometric fit, for both right and left side detections.* ***c*** *Median reaction times for each rewarded condition for threshold and maximum levels of saliency (right and left sides pooled together). Boxplots show the median and interquartile range (box limits) and 1.5 x interquartile range (whiskers). Visual reaction times*

*were significantly shorter for UST compared to MST (ANOVA, n=1395 trials, $F(1,1391)= 54.075$, $p=3.28x10^{-13}$, corrected for 2 multiple comparisons (Bonferroni-Holm)). *: p<0.05; ***: p<0.001). Note that tactile and visual reaction times were similar, ruling out the possibility of a sequential detection strategy where one modality would be sampled before the other one.* ***d*** *D-prime of contralateral tactile detection at threshold saliency. V1 silencing did not affect tactile performance (ANOVA, Early silencing: n=6 sessions, $F(1,12)=3.78$, $p=0.1513$; Late silencing: n=9 sessions, $F(1,14)=0.12$, $p=0.73$). Thus, for MST, silencing late V1 activity impaired visual but not tactile detection, indicating that late activity per se is not required for licking behavior.* ***e*** *Left: Psychometric curves for experiments with left hemisphere V1 silencing (same experiments as Fig. 6c), for visual-only trials. Data points for MST Max left are missing because these were not assessed in the experimental protocol (same for Max left and Thr left conditions in control experiments). Since monocular stimuli were used and left hemisphere V1 was silenced, potential effects were expected for stimuli on the right side (contralateral) but not on the left side (ipsilateral). (ANOVA, n=7 sessions, UST Early, $F(1,10.7)=39$, $p=0.0002$; UST Late, $F(1,10.7)$, $p=0.0178$; MST Early, $F(1,12)=32$, $p=0.0002$; MST Late, $F(1,18)=38$, $p=3x10^{-5}$). *: p<0.05. Note that late silencing for UST mice had a significant effect on the percentage of right choices, but no effect on the corresponding d-prime. Right: psychometric curves for control experiments where the optic fiber was placed above the mouse headbar cement and therefore not above V1 (all not significant). Errorbars denote interquartile range.* ***f*** *Average z-scored activity for all recorded left-hemisphere V1 neurons for each of all 9 possible visual-only stimulus-response combinations for UST (left) and MST (right) trained mice. Neurons for which no more than three trials were present in the given condition were omitted. For each condition, neurons were sorted according to their mean z-score between 50 and 500 ms.* ***g*** *Average z-scored firing rate of responsive V1 neurons during visual-only trials, for three different conditions eliciting the same licking response (left: contralateral lick, right: ipsilateral lick), showing that late V1 activity cannot be explained by licking alone (see also Supplementary Fig. 3). Activity was aligned to the first lick in the response window. Licks made to the wrong side were termed "errors". Same neurons as in Fig. 6c. Shaded area: bootstrapped 95% confidence intervals.* ***h*** *Average z-scored firing rate of responsive V1 neurons during contralateral tactile trials, split by choice, for UST and MST mice. For UST, surrogate hits correspond to licks to the same side of the tactile stimulus (although unrewarded) and surrogate misses correspond to trials without licks. Late activity was present only in MST Hits, indicating that the same stimulus and the same behavioral response triggered late activity in a context-dependent manner. Shaded area: bootstrapped 95% confidence intervals.*



Supplementary Figure 3.9 Noise correlations in auditory trials and latency of the drop in noise correlations during visual trials in task A. ***a*** *Decrease in noise correlation (NC) of V1 cell pairs with respect to baseline for auditory hits and misses across cohorts (same conventions as Fig. 5c). For UST mice there were too few auditory hits to compute noise correlations. In MST mice, noise correlations decreased during hits (ANOVA, n=14462 trials, $F(1,28847) = 99.90$, $p = 1.7*10^{-23}$) and increased during misses (ANOVA, n=13656 trials, $F(1,27395) = 22.61$, $p=1.99*10^{-6}$). The fact that NCs also decrease in auditory hits indicates that the drop in NCs in the visual cortex is not specific to visual trials, and possibly is a more general mechanism that may subserve decision making. Boxplots show the median and interquartile range (box limits) and 1 x interquartile range (whiskers).* ***b*** *Drop in NCs during visual trials, but relative to the first lick for each tertile of reaction times for UST and MST. Same as Fig. 5f, but aligned to reaction time. No significant correlation is found (p=0.09), in contrast with Fig. 5f, indicating that the drop in NCs precedes reaction time by a relatively constant time lag.* ***c*** *Noise correlations (NC) of V1 cell pairs over time aligned to the first lick for the different trial types and cohorts. Dotted line shows the threshold for a significant drop in NCs with respect to baseline (-1000 to -500 ms relative to first lick). Noise correlations decreased most for visual hits in visually trained mice (UST and MST), but not in NE mice. For UST mice there were too few auditory hits to compute noise correlations. Shaded area corresponds to s.e.m.*

Chapter 4: Triple dissociation of visual, auditory and motor processing in primary visual cortex

Matthijs N. Oude Lohuis
Pietro Marchesi
Umberto Olcese*
Cyriel Pennartz*

*These authors contributed equally

Manuscript submitted, preprint available on BioRxiv:
https://doi.org/10.1101/2022.06.29.498156

Abstract

Primary sensory cortices respond to crossmodal stimuli, for example auditory responses are found in primary visual cortex (V1). However, it remains unclear whether these responses reflect sensory inputs or behavioural modulation through sound-evoked body movement. We address this controversy by showing that sound-evoked activity in V1 of awake mice can be dissociated into auditory and behavioural components with distinct spatiotemporal profiles. The auditory component began at ~27 ms, was found in superficial and deep layers and originated from auditory cortex, as shown by inactivation by muscimol. Sound-evoked orofacial movements correlated with V1 neural activity starting at ~80-100 ms, was more sustained and explained auditory frequency-tuning. Visual, auditory and motor activity were expressed by segregated neuronal populations. During simultaneous audiovisual stimulation, visual representations remained dissociable from auditory and motor-related activity. This threefold dissociability of auditory, motor and visual processing is central to understanding how distinct inputs to visual cortex interact to support vision.

Introduction

During our everyday lives, we sample the world through active exploration with our different senses. Already in primary sensory cortices, contextual signals about ongoing events in other sensory and motor modalities are integrated with modality-specific signals to enable meaningful sensory processing (Ghazanfar and Schroeder, 2006; Guitchounts et al., 2020; Jones and Powell, 1970; Kayser and Logothetis, 2007; Meijer et al., 2019; Niell and Stryker, 2010; Stringer et al., 2019). In general, integration of crossmodal and motor information may enhance sensory detection and discrimination by way of Bayesian cue integration, may subserve cross-modal predictions, and may underlie the contextual and modally distinct representations characteristic of conscious experience (Fetsch et al., 2013; Pennartz, 2015, 2009). In rodents, primary and secondary visual and auditory cortices share direct anatomical connections (Budinger and Scheich, 2009; Budinger et al., 2006; Campi et al., 2010; Cappe and Barone, 2005; Falchier et al., 2002; Miller and Vogt, 1984; Paperna and Malach, 1991; Rockland and Ojima, 2003) and auditory inputs to primary visual cortex (V1) have been found to target L1 and L5/L6 (Ibrahim et al., 2016; Iurilli et al., 2012; Mesik et al., 2019; Rockland and Ojima, 2003). Auditory inputs affect visual response properties such as orientation and contrast tuning in V1 (Ibrahim et al., 2016; Meijer et al., 2017), with some V1 neurons directly responding to sounds (Meijer et al., 2017). Even more strikingly, selectivity to auditory features such as spatial location and frequency has been reported in cat visual cortex (Fishman and Michael, 1973; Morrell, 1972; Spinelli et al., 1968) and mouse V1 (Knöpfel et al., 2019).

In addition to crossmodal inputs, other factors strongly impact activity in primary sensory areas. Locomotion and increases in arousal both desynchronize spike patterns and increase V1 activity (Fu et al., 2014; Niell and Stryker, 2010; Vinck et al., 2015c). Also eye, head and orofacial movements lead to marked activity changes (Bouvier et al., 2020; Guitchounts et al., 2020; Stringer et al., 2019). The exact function of motor signaling to primary sensory cortices remains unknown, but it possibly provides an efference copy serving to predict the sensory consequences of body movement. Failing to account for these motor-related influences risks misinterpreting the observed sensory cortical activity (Musall et al., 2019; Zagha et al., 2022). Indeed, V1 activity during sound clips has been argued to have a behavioral rather than sensory origin and relate to stereotyped sound-evoked movements (Bimbard et al., 2021). The origins of crossmodal activity in sensory cortex are thus unclear and dissociating auditory from behavioral signals in V1 is central to correctly interpret (multi-)sensory activity and to understand how distinct inputs to visual cortex interact to support vision.

To examine whether visual, auditory and motor processing can be dissociated in V1, we trained mice on a task which required them not only to detect sensory changes, but also to distinguish or segregate the modality in which the change occurred. In awake animals performing this task, auditory stimuli evoke frequency-tuned neuronal responses in V1, that are largely, but not completely,

explained by orofacial movements. We disentangled auditory-related from motor-related activity using multi-area recordings, task manipulations, pharmacological interventions and optogenetics. An early, sound-evoked component of V1 responses originates from auditory cortex, is transient, and is found predominantly in superficial and deep layers. In contrast, motor-related activity following auditory stimuli results from rapid orofacial movements with an onset-latency of 60-100 ms, is found mostly in superficial layers, and underlies the bulk of late sound-evoked neural activity changes in visual cortex, but not in auditory cortex. Jointly, these signals strongly affected visual cortical activity in a way that leaves visual stimulus coding intact.

Results

To disentangle visual, auditory and behavior-related signaling in V1 we presented three different cohorts of head-fixed mice with the same sensory stimuli, but trained them to report only visual stimuli, both auditory and visual stimuli, or neither (Fig. 4.1a). Visual stimuli were continuous full-field drifting gratings and visual trials consisted of uncued, occasional orientation changes. Auditory stimuli consisted of weighted combinations of five harmonic tones that occasionally changed frequency (auditory trial; Ext. Data Fig. 4.1a-c). Noncontingently exposed mice (NE; N=7) were not rewarded for licking after any stimulus change, but were pseudorandomly rewarded for spontaneous licks and therefore served as naive control animals. Unisensory trained mice (UST; N=4) were only rewarded for lick responses to visual changes and learned to ignore auditory changes. In the multisensory task version (MST; N=17), responses to both visual and auditory changes were rewarded and animals were trained to lick left to auditory changes and right to visual changes (or vice versa). Thus, not only the change but also modality identity was behaviorally reported through directional lick responses. Hit rate scaled with stimulus saliency (amount of visual or auditory feature change) only in modalities for which mice were rewarded (Fig. 4.1b), confirming effective manipulation of the task-relevance of sensory changes.

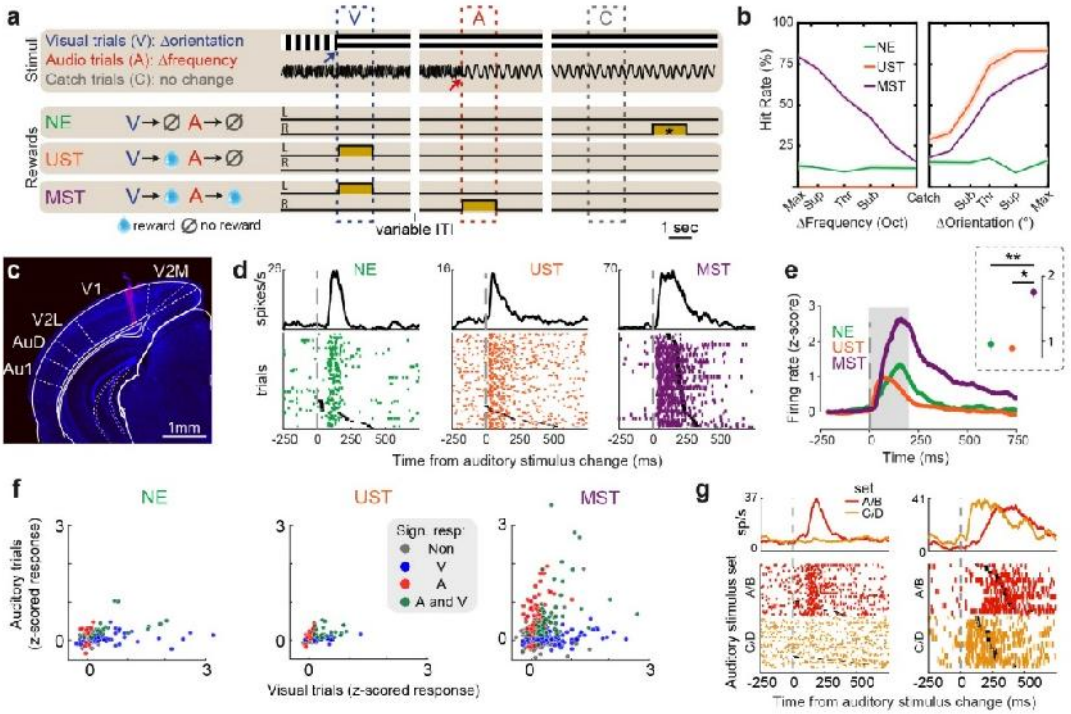


Figure 4.1: Task dependent recruitment of auditory activity in visual cortex upon behavioral relevance. a) Stimuli and reward contingencies for the three cohorts of mice. Visual and auditory stimuli were continuously presented and visual trials (V) consisted of uncued orientation changes (blue arrow), auditory trials (A) of uncued frequency changes (red arrow), and catch trials (C) of no change. Noncontingently exposed mice (NE) were not rewarded for licking after stimulus changes, unisensory trained mice (UST) for licks after visual but not auditory changes, and multisensory trained mice (MST) for reporting both visual and auditory changes. In MST mice, each lick spout was associated with a single modality. Ochre rectangle indicates reward availability. *For NE mice, reward windows were temporally decorrelated from the sensory stimuli, and randomly occurring outside the stimulation period. The dotted trial windows indicate the time window used post-hoc to compare stimulus-related lick rates across cohorts. ITI: inter-trial interval. **b)** Average hit rates at different stimulus saliencies for each cohort. Sub: subthreshold, thr: threshold, sup: suprathreshold, max: maximal). Behavioral hit rates increased as a function of the amount of auditory change (step size in frequency) in MST mice (left panel), and as a function of visual orientation change for UST and MST mice (right panel), in line with their reward schedule. UST mice were only rewarded for licks to the visual lick spout and therefore chose not to lick the auditory lick spout. Mean ± SEM. **c)** DAPI-stained (blue) coronal section showing electrode track in left V1 stained with DiI (red) ±3.56 mm posterior to Bregma. V2M and V2L: medial and lateral secondary visual cortex, respectively. Au1: primary auditory cortex. AuD: dorsal secondary auditory cortex. **d)** Raster plots showing sound-evoked activity in example V1 neurons from each cohort. Black ticks indicate the first lick after stimulus change, which was rewarded only in MST mice. **e)** Averaged z-scored firing rate (referenced to baseline) for auditory responsive V1 neurons to preferred post-change auditory stimulus. Inset shows response averaged across the shaded analysis window (0-200 ms). *p<0.05, **p<0.01. Mean ± SEM. **f)** Scatter plot of z-scored firing rate, corrected for baseline activity (analysis window 0-200 ms) and following auditory and visual stimulus changes. Each dot is a neuron. (NE: N=163; UST: N=128; MST: N=812). Colors denote cells with a significant response to any modal input (gray: no significant response). **g)** Individual V1 neurons show frequency-specific sound-evoked activity. Raster plots as

in (d), but for the two sets of auditory post-change frequencies (where A and B are two similar frequencies and C and D as well; Ext. Data Fig. 4.1e). Left example from NE mouse, right from MST mouse.

Sound-evoked activity in primary visual cortex

We recorded single unit and LFP activity in V1 (Fig. 4.1c). During recording sessions, two levels of change saliency were presented per modality: threshold level (Vthr and Athr, individually titrated per animal based on psychophysical data; see Methods) and maximal saliency (Vmax and Amax, 90° and ½ octave change, respectively; Ext. Data. Fig. 4.1d). Furthermore, stimuli were restricted to four orientations and four frequencies. Within these visual or auditory stimulus sets, stimuli A and B were highly similar (differing at threshold level), as were C and D, and nearby stimuli were grouped into pairs for analyses (set A/B vs. set C/D; see Ext. Data Fig. 4.1e).

Auditory frequency changes at maximal saliency (Amax) induced spiking activity in V1 neurons of mice from all three cohorts (Fig. 4.1d). These responses were generally transient and subsided after 200 ms. A sizeable fraction of V1 neurons had a significantly higher firing rate (0-200 ms) compared to baseline following at least one of the post-change auditory stimuli (Wilcoxon signed rank test, $p<0.05$; during Amax trials). This fraction was similar across cohorts (NE: 25.5 ± 4.0%, n=5 sessions; UST: 34.8 ± 4.8% n=4 sessions; MST: 20.9 ± 0.7% n=27 sessions; $p=0.30$, Kruskal-Wallis test). A large fraction of V1 neurons also responded to one of the two visual orientation changes in all three cohorts (NE: 42.4 ± 7.9%; UST: 53.2 ± 6.8 %; MST: 30.2 ± 0.8%; mean ± SEM; $p=0.13$, same sessions as above, Kruskal-Wallis test). To quantify and compare neuronal activity, firing rates during auditory and visual trials were z-scored (Fig. 4.1e,f). Auditory responses of V1 neurons (0-200 ms after stimulus change) were particularly strong in MST animals, in which responses exceeded those observed in NE and UST animals ($F(2,232)=4.96$, $p=0.01$; Posthoc comparison: MST vs NE: $F(1,229)=4.40$, $p=0.037$; NE vs UST: $F(1,229)=0.02$, $p=0.876$; MST vs UST: $F(1,229)=7.75$, $p=0.006$). V1 neurons were previously found to be selective for auditory stimulus features such as sound frequency (Fishman and Michael, 1973; Knöpfel et al., 2019; Morrell, 1972; Spinelli et al., 1968). Indeed, individual neurons displayed stimulus-specific sound-evoked firing rate responses (Fig. 4.1g) and a surprisingly large fraction of V1 neurons discriminated post-change auditory stimulus identity (12.7% ± 3.7%, Mean +- SEM across sessions, A/B vs C/D Amax trials, permutation test; $p<0.05$), which was comparable across cohorts (NE: 11.9 ± 2.7%; UST: 16.1 ± 4.5%; MST: 12.1 ± 1.9%; $p=0.68$, Kruskal-Wallis test). This was significantly smaller than the fraction of orientation-selective cells across cohorts (31.9 ± 3.7%; fraction frequency- versus orientation-tuned: $F(1,52)=18.48$, $p=7.54*10^{-5}$). This fraction is lower than commonly reported in the literature, presumably because we only sampled four irregularly spaced grating orientations (Ext. Data Fig. 4.1e).

Motor contribution to sound-evoked activity in visual cortex

The enhanced sound-evoked responses in V1 of MST animals are potentially explained by licking as only these mice were rewarded for reporting auditory changes. Auditory lick reaction times in MST mice were around 320 ms (median, IQR: 219-446 ms) and faster than visual reaction times (median: 404 ms; IQR: 320-509 ms; $F(1,270) = 105.65$, $p=3.94 * 10^{-21}$), suggesting motor-related activity changes in V1 could already play a role during early time windows after the change in sound (0-200 ms). We recorded the mice's faces and often observed orofacial movements following auditory stimulus changes in trained, but also untrained mice. We therefore explored to what extent sound-evoked activity in V1 was related to (stereotypical) movements (Bimbard et al., 2021; Williams et al., 2021). To investigate this, motion energy was extracted from video footage (video ME) as the overall pixel intensity difference between consecutive frames (Fig. 4.2a), which included snout, whisking and licking movements.

Visual stimulus changes induced movements only in UST and MST mice, the cohorts rewarded for licking. A change in auditory stimulus was followed by strong movements in MST mice, but to a lesser extent also in NE and UST mice (Fig. 4.2b). In Amax auditory trials, these orofacial movements started roughly 60-100 ms after the stimulus change, but our temporal resolution was limited by a frame rate of 25 fps. In MST animals, instrumental licking movements dominated, while in NE and UST animals, where sounds were behaviorally irrelevant, they evoked mainly whisking and snout movements (Ext. Data Fig. 4.2a,b). Similarly, if we removed the lick spout in MST mice, auditory stimuli continued to evoke orofacial movements other than licking, while movements disappeared altogether following visual stimuli (Ext. Data Fig. 4.2c-e). Sounds thus evoked orofacial movements whether relevant or irrelevant to the current task, in contrast to visual stimuli, which only led to reward-contingent movements.

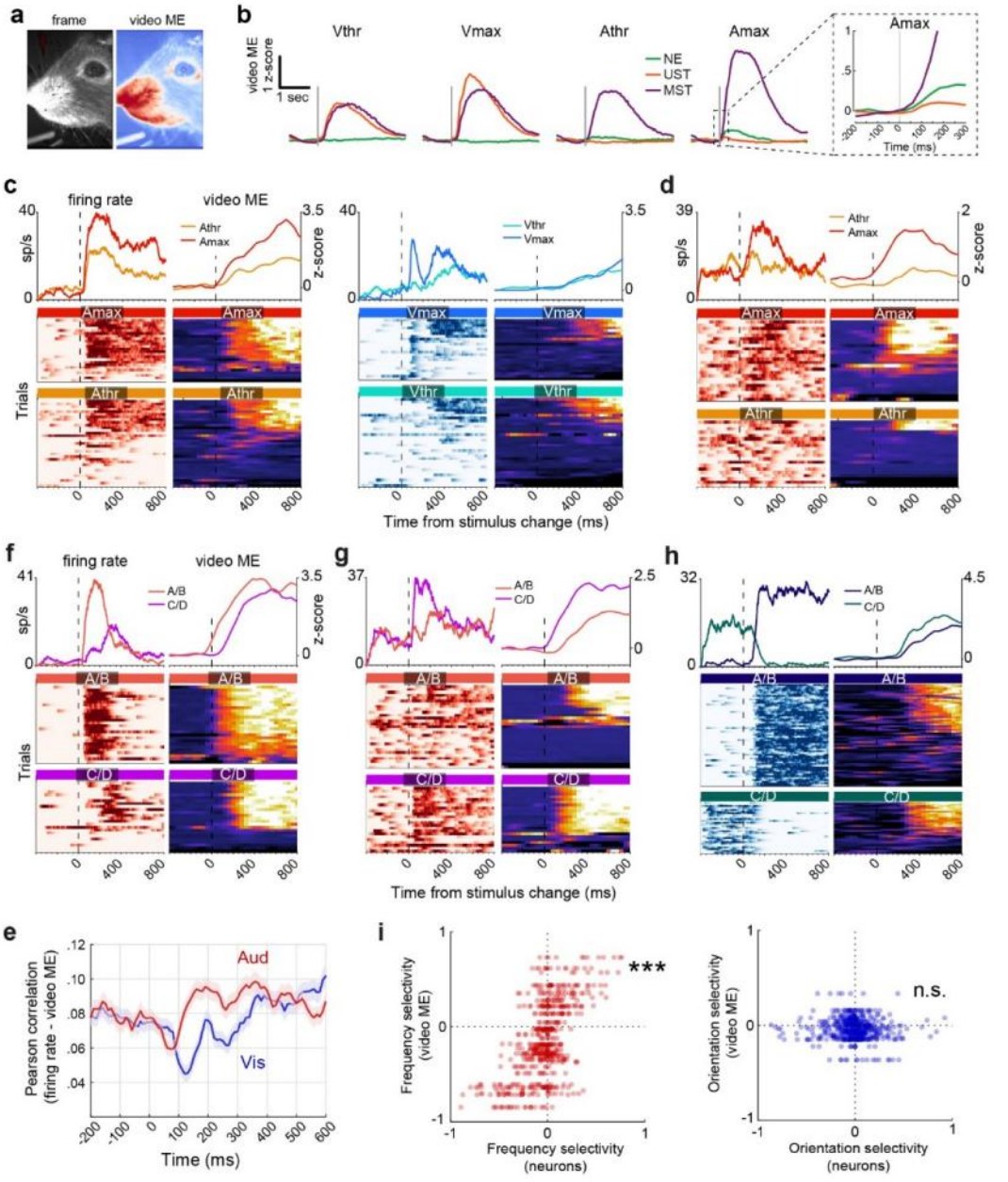


Figure 4.2 – Sound-evoked orofacial movements explain frequency-tuned activity in visual cortex. a) Example frame and motion energy from the video (video ME). Red indicates more orofacial movements. **b)** Video ME for auditory (threshold and maximal saliency, Athr and Amax) and visual (Vthr and Vmax) trials across cohorts. Inset shows a zoom in of boxed region in Amax trials. **c)** Trial-by-trial spiking activity correlates with video ME in an example V1 neuron (MST animal). Columns from left to right: firing rate during auditory trials (red is high firing rate), video ME during auditory trials (warm colors is increased orofacial motion), firing rate during visual trials (blue is high firing

rate), video ME during visual trials. Trials include hits and misses and are separated by saliency (thr and max) and sorted by magnitude of video ME. Upper panels show trial-averages of heatmaps below. Heatmap color range is the axis range of the upper plot. **d)** Same as c, but for another example V1 neuron (MST cohort) with sound-evoked spiking activity in Amax trials without changes in video ME in a subset of trials. **e)** Average Pearson correlation of single neuron firing rate to video ME during auditory (red) and visual (blue) trials over time. **f-h)** Same as *c*, but with trials separated based on visual and auditory post-change stimulus class (set A/B versus set C/D). In the left two examples, both firing rate and video ME discriminate post-change auditory feature, and firing rate and video ME are correlated across trials. In the right example, firing rate discriminated grating orientation independently of orofacial movements and video ME was largely unrelated to grating orientation. Pre-change coding resulted from the fact that gratings changed from A/B to C/D, or vice versa. Neurons from MST mice. **i)** Left: frequency selectivity of V1 neurons (x-axis) and video ME-based selectivity in the same session (y-axis). Selectivity was measured as AUC, rescaled between -1 and 1. Each dot is a neuron (all cohorts combined). Right: same but for orientation selectivity.

Sound-evoked movements underlie frequency tuning

For of a number of individual V1 neurons, activity correlated with video ME (Ext. Data Fig. 4.3a) (Stringer et al., 2019). Also during auditory trials, some neurons showed an overall similarity between trial-to-trial firing rate and video ME (Fig. 4.2c). However, example neurons also showed sound-evoked activity in trials without increases in video ME (Fig. 4.2d). The increase in orofacial movements after auditory changes could therefore underlie sound-evoked activity in V1 at least in a fraction of neurons. As a first test of this hypothesis, we correlated firing rate with video ME during auditory and visual trials for all neurons (Fig. 4.2e). Relative to baseline, the average correlation decreased for visual trials, likely due to visually driven activity unrelated to ongoing movements. Following auditory changes, the average correlation transiently dropped and then increased after roughly 100 ms, in line with the observed movement-related activity of individual V1 neurons.

We next wondered whether movements could also underlie the observed auditory frequency-specific V1 activity (Fig. 4.1g). As illustrated for two example neurons (Fig. 4.2f,g), frequency-tuned activity was strongly aligned to variability in video ME in those trials. Both neuronal activity and video ME thus responded to auditory changes in a stimulus-specific manner. This was not the case for orientation-tuning (Fig. 4.2h). Auditory feature tuning could therefore result from 'motor-tuning'. Indeed, across the full population, tuning to either one of the auditory stimuli (quantified using the receiver operating characteristic, ROC) was accompanied by strong movements to that stimulus ($F(1,688)=286.27$, $R=0.542$, $p=5.92*10^{-54}$), but this was not the case for visual neuronal selectivity ($F(1,692)=2.90$, $R=0.065$, $p=0.09$; Fig. 4.2i). Furthermore, if strong motor activity drove frequency selectivity, one would expect all simultaneously recorded cells to be preferentially tuned to the same auditory stimulus (viz. that elicited most movement). Indeed, simultaneously recorded V1 neurons responded to the same auditory stimuli, while their preferred grating orientation was mixed (Ext. Data Fig. 4.3b). These findings were corroborated in a separate set of animals trained on a stimulus detection (rather than change detection) version of the task, which allowed us to test a more extended range of visual and auditory stimuli (see Methods; Ext. Data Fig. 4.4). We found similar results when training population decoders on V1 population spiking

activity or on high-dimensional video data (Ext. Data Fig. 4.3c-k; dimensions are understood here as different PCA components; Stringer et al. 2019). Auditory frequency could be decoded from neural and video data in a highly correlated manner, whereas grating orientation could be deduced only from V1 activity. In sum, auditory stimuli led to fast orofacial movements that correlated with sound-induced and tone-specific V1 activity.

Dissociable auditory and motor-related signals

A purely behavioral origin of sound-evoked activity in V1 would be in disagreement with some previous findings. Primary auditory and visual cortices are monosynaptically connected and sound-evoked activity has been reported under anesthetized conditions where overt movements and changes in behavioral state play a very minor role, if any (Henschke et al., 2015; Ibrahim et al., 2016; Iurilli et al., 2012). Furthermore, in our recordings, many neurons showed short-latency sound-evoked activity also in trials without apparent changes in motor activity (e.g. Fig. 4.2d). We therefore explored whether auditory activity can be disentangled from motor-related activity in visual cortex.

First, we compared sound-evoked V1 activity to spiking activity in auditory cortex, where the importance of bottom-up auditory signaling is well established (AC; including primary auditory cortex, anterior auditory field and dorsoposterior auditory cortex; Fig. 4.3a). AC strongly responded to auditory stimulus changes, but only minimally to visual orientation changes (Fig. 4.3b, *right*). We aggregated population activity across neurons to achieve a higher temporal resolution than single-unit activity and compared the latency to firing-rate increases evoked by visual and auditory stimuli in V1 and AC (Fig. 4.3b). The auditory response started at 18 ms in AC and 27 ms in V1. The small increase in AC activity after visual stimuli became significant only after 156 ms. This short onset latency of sound-evoked activity in visual cortex fits auditory-related inputs to V1. Visual responses started at 54 ms in V1, matching the canonical retinogeniculate drive (Schnabel et al., 2018).

To further disentangle auditory and motor-related signals, we built a regression model (Fig. 4.3c) to predict single-neuron activity based on trial number (e.g. to account for drift in motivational state), visual and auditory stimulus features and motor activity (the first 25 video PCs). Given that video predictors represented orofacial movement, we termed the correspondingly explained neural activity *motor-related*. This approach succeeded in separating the temporally overlapping contributions of sensory stimuli and motor-related activity to trial-by- trial single-neuron firing rate (Ext. Data Fig. 4.5a). Visual stimuli explained most V1 variance during 0-200 ms post-stimulus change, followed by motor-related activity as next best predictor (Fig. 4.3d). Auditory stimuli explained a modest but significant fraction of the variance. Motor activity explained roughly three times as much variance as the auditory stimuli. In auditory cortex, firing rate was best explained by auditory stimuli, with only a minor contribution of trial-number and motor

activity, and no apparent influence of visual stimuli. In V1, visual encoding was found largely in a separate set of neurons (Fig. 4.3e) and visual EV was not correlated to motor EV (r=-0.03, $F(1,471)=0.61$, p=0.44), nor auditory EV (r=-0.08, $F(1,484)=3.49$, p=0.06). Neurons were observed that uniquely encoded auditory or motor features, as well as jointly auditory-motor coding neurons and the overall EV was moderately correlated (R=0.32, $F(1,473)=40.56$, $p=4.52x10^{-10}$). To examine how auditory or motor-related activity distinctly contributed to the average response in V1, single neuron firing rate was predicted using only subsets of predictors in our model (Fig. 4.3f). Averaging this predicted activity across V1 neurons revealed that activity during visual trials was mostly visual in the early phase (0-200 ms), with a late motor-related component (>200 ms). Activity during auditory trials presented a combination of temporally overlapping auditory and motor-related components, with only early activity (<100 ms) being predominantly auditory. This was strikingly not the case in auditory cortex, which showed a negligible contribution of motor activity to the averaged response. The sound-evoked activity in V1 was thus composed of a distinct, early auditory and a later motor-related component. Performing these analyses for the different cohorts showed that auditory-related activity in V1 was similar across cohorts and that the larger sound-evoked response in MST mice (Fig. 4.1e) was likely the result of increased motor-related activity due to instrumental licking in this cohort (Ext. Data Fig. 4.5b,c).

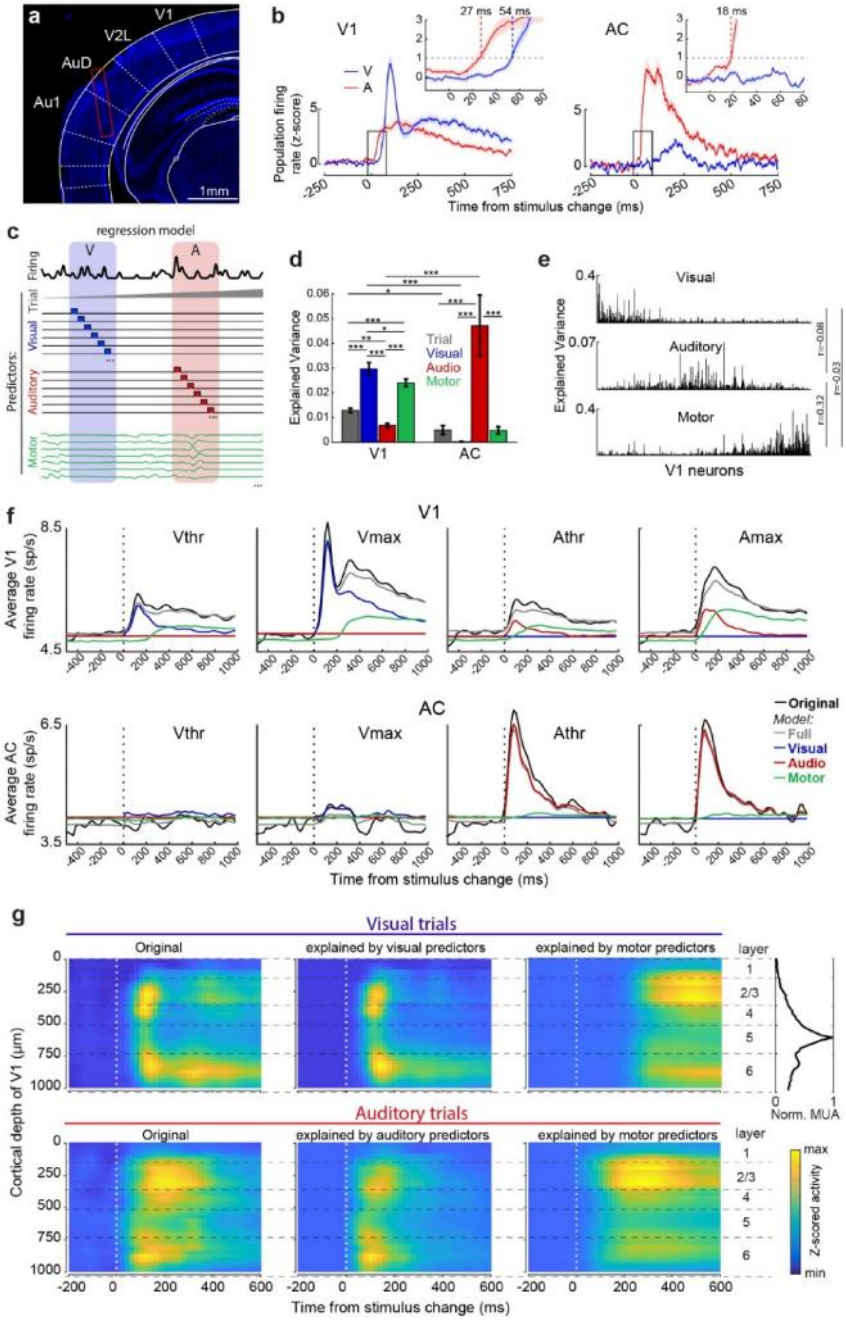


Figure 4.3 – Temporal and spatial dissociation of auditory and motor-related activity in auditory and visual cortex. a) DAPI-stained (blue) coronal section showing electrode track in Auditory dorsal (AuD) and primary Auditory cortex (Au1) approximately ±2.46 mm posterior to Bregma. **b)** Population firing rate after auditory and visual stimulus changes in V1 (left) and AC (right). Inset

shows close-up of boxed region and the latency to cross a threshold of Z-score > 1. Line and shaded region are mean and SEM across sessions. **c)** Schematic of regression model. Single neuron firing rate was predicted as a linear combination of four sets of predictors. A 'Trial' predictor captured rate fluctuation throughout the session. For visual (V) and auditory (A) stimuli a set of separate predictors spanned one second after stimulus change per saliency level and feature (orientation and auditory frequency; shown only for one trial type) to capture neuron-specific response patterns. Motor predictors included the top 25 video PCs. **d)** Explained variance during 0-200 ms of all trials for each category of predictors. Mean ± SEM across neurons. Results based on 51 sessions; NE: 9, UST: 10, MST: 32 sessions, 19.451 trials, 790 V1 and 99 AC neurons. Asterisks indicate significance of a post hoc F-test on the relevant contrast in the linear mixed effects model. *p<0.05, **p<0.01, ***p<0.001. **e)** Distribution of explained variance across V1 neurons for each category of predictors (except Trial), showing encoding of sensory and motor variables in distinct and overlapping neurons. Neurons are sorted based on visual minus motor encoding and centered by auditory encoding. **f)** The firing rate of single neurons was predicted using only subsets of predictors (visual: blue, auditory: red, motor: green) or all predictors (gray) and compared to the original firing rate (black). Shown is the firing rate during different trial types averaged across V1 neurons (upper row) or AC neurons (bottom row). **g)** Heatmap of z-scored firing rate with V1 neurons binned by cortical depth. Leftmost heatmaps show original data; middle and right panels show predictions using sensory or motor variables from the regression model. Rightmost panel shows normalized multi-unit activity peaking in L5 used to demarcate layers (see Ext. Data Fig. 6). Throughout the figure neurons from all cohorts were combined. See Ext. Data Fig. 5b,c for cohort-specific findings.

Laminar organization of auditory and motor-related inputs to V1

As V1 was recorded with microelectrode arrays that spanned the different layers, we next wondered whether the auditory and motor-related components had distinct spatiotemporal profiles. The electrode position was aligned to the cortical depth for each V1 penetration (Ext. Data Fig. 4.6, Methods). We constructed heatmaps of z-scored firing rate as a function of cortical depth and time relative to the stimulus change, using either the raw data or the activity predicted from stimulus or motor components only (Fig. 4.3g). Visually induced spiking activity clustered in layer 4 (L4) to L2/3 as well as later in L6, while motor-related activity accounted for the later component predominantly in superficial layers 2/3, but also L6. Sound-evoked activity was decomposed into an early auditory-related component spanning the layers, but most prominently in L2/3 and L5/6, with motor-related activity again mostly contributing in superficial layers. Note how motor activity explained a similar pattern during auditory and visual trials, but shifted to an earlier time window in auditory as compared to visual trials. These results were corroborated by analyses of the local field potential, showing similar onset latencies of auditory and visual evoked activity and predominance of motor-related activity in superficial layers (Ext. Data Fig. 4.7). When estimating the laminar organization of visual and auditory components with a model-free approach and plotting the onset latency of visual- and auditory-evoked single-unit activity as a function of cortical depth, we found similar results as for the GLM analysis of firing patterns (Ext. Data Fig. 4.7h). Together, these results indicate different spatiotemporal profiles for visual, auditory, and motor-related components in V1. Our data are most consistent with auditory stimuli evoking fast auditory-related inputs predominantly to deep layers and later, motor-related activity mostly in superficial layers.

Auditory cortex as a source of early sound-evoked activity in V1

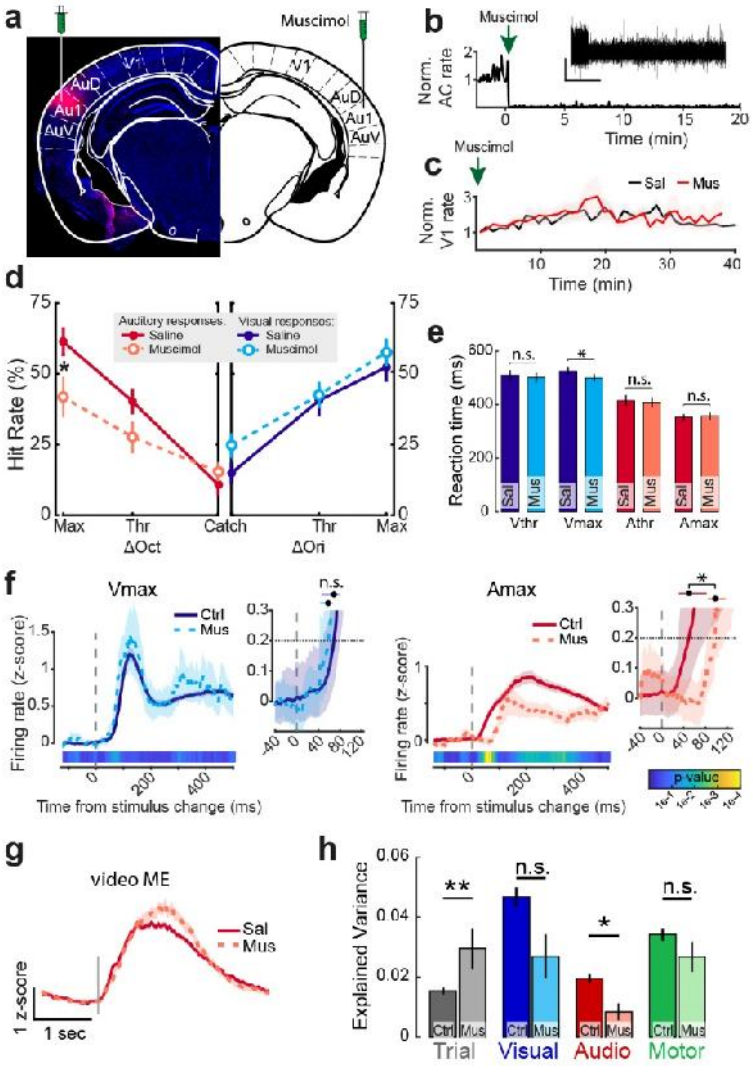
The short-onset, sound-evoked activity in V1 (27 ms), shortly after AC (19 ms, Fig. 4.3b), suggests that AC may be a source of early auditory-related activity in V1. To test this, we first bilaterally injected AAV-CamKIIa-ChR2-eYFP in AC of naive mice (Ext. Data Fig. 4.8a). Axonal terminals were observed in layers L1 and L5/6 of V1 (Ext. Data Fig. 4.8b,c). This AC-V1 projection pattern matches that found in earlier studies (Ibrahim et al., 2016; Rockland and Ojima, 2003). In a subset of animals, cell bodies were photostimulated in AC and the LFP response was recorded in V1 (Ext. Data Fig. 4.8f-j). Ten millisecond laser pulses over AC led to short-latency LFP deflections in V1, indicative of AC to V1 connections (Ext. Data Fig. 4.8g), in line with (Ibrahim et al., 2016; Iurilli et al., 2012).

Next, we bilaterally injected AC with muscimol in MST mice during task performance (Fig. 4.4a). Muscimol infusion immediately abolished spontaneous multi-unit activity in AC (Fig. 4.4b) but did not affect V1 firing rate as compared to the saline control (Fig. 4.4c). Muscimol moderately impaired auditory hit rates at maximal saliency (Amax: $F(1,19)=6.41$, $p=0.020$; catch: Athr: $F(1,19)=4.26$, $p=0.053$; $F(1,19)=1.69$, $p=0.210$; n=11 saline, 11 muscimol sessions; Fig. 4.4d) without effects on visual hit rates (catch: $F(1,19)=3.88$, $p=0.064$; Vthr: $F(1,22)=0.09$, $p=0.773$; Vmax: $F(1,18)=0.43$, $p=0.522$). Reaction times were unaffected, except for a small reduction in Vmax trials (Vthr: $F(1,569)=0.16$, $p=0.686$; Vmax: $F(1,748)=6.03$, $p=0.014$; Athr: $F(1,487)=0.05$, $p=0.818$; Amax: $F(1,705)=0.30$, $p=0.581$; Fig. 4.4e).

Recordings during task performance revealed that the firing rate of V1 neurons following visual stimuli during AC inactivation was similar to control experiments (Fig. 4.4f). However, AC inactivation affected V1 firing during auditory trials particularly in the first 100 ms and was still associated with a later increase in average firing rate. The onset latency of increased firing was not significantly different for visual stimuli (control: 69 ms (47 - 81 ms) versus muscimol: 58 ms (43 - 67) ms; median and bootstrapped 95% CI; maximal saliency), but delayed for auditory stimuli (control: 50 ms (30 - 82) ms versus muscimol: 98 ms (84 - 117) ms; bootstrap test; $p<0.05$). Thus, muscimol suppressed the early component of the V1 response to auditory stimuli (from about 30 to 80 ms; Fig. 4.4f). Muscimol did not affect activity during Athr trials consistent with the fact that V1 activity

during these trials was mostly motor-related, with only a minimal auditory-related component (see Fig. 4.3f).

Figure 4.4: Muscimol in auditory cortex impairs auditory change detection and early evoked activity in visual cortex. a) Histological section and schematic of approach to bilaterally inactivate auditory cortex with muscimol. Red blob shows BODIPY TMR-X conjugated muscimol localized mainly in primary auditory cortex (Au1). AuD/AuV: dorsal/ventral secondary auditory cortex. **b)** Multi-unit activity in AC normalized to session start before and after muscimol injection, showing a severe reduction of spiking activity. Inset shows a high-pass voltage trace (cut off at 500 Hz) at one electrode during the same timeframe. Scale bar indicates 100 microvolt and 5 minutes. **c)** Average single-unit activity in V1 normalized to session start for saline and muscimol sessions. Injections were 1-5 minutes prior to recordings. Line and shading indicate mean ± SEM across sessions. **d)** Hit rates in



auditory trials (left panel) and visual trials (right panel) for muscimol and saline sessions. Mean ± SEM across sessions. *p<0.05. **e)** Reaction times. Mean ± SEM across trials. *p<0.05. **f)** V1 spiking

activity with and without AC inactivation during visual (left) and auditory (right) trials of maximal saliency. Main panels show average spiking activity with the line and shading indicating mean ± SEM across neurons. The lower colormap scales with p-value of bootstrapped test of difference per time bin (n=10.000 bootstraps, $p<0.05$, Bonferroni corrected; control neurons were subsampled to match the number of recorded V1 neurons during AC muscimol infusion; dataset: 32 control sessions, 570 V1 neurons; 4 muscimol sessions, 53 V1 neurons). The insets show a close up of the early post-stimulus time window and bootstrapped activity (line and shading indicate mean ± 95% confidence interval). The dashed line indicates statistical threshold of deviations from baseline. Upper horizontal dot and error bar indicate when bootstrapped activity significantly exceeded baseline activity. The non-overlap of these distributions showed a significant difference in onset latency ($p<0.05$). **g)** Overall orofacial motion energy during auditory trials. Grey line: stimulus change. Mean ± SEM across trials. **h)** Explained variance of V1 spiking activity by each predictor subset in the regression model with and without AC inactivation. Mean ± SEM across neurons. *$p<0.05$. **$p<0.01$.

Given that auditory performance was reduced but not abolished by muscimol injection, the residual auditory hits were still associated with instrumental licking and orofacial movements (Fig. 4.4g), which could underlie late V1 firing during auditory trials. We tested whether muscimol selectively affected the contribution of auditory predictors to V1 firing rate, while preserving other components (Fig. 4.4h). Auditory predictors explained significantly less variance under muscimol ($p=0.04$; Wilcoxon rank sum test), while visual and motor predictors were not associated with significant reductions (Vis: $p=0.23$; Motor: $p=0.95$). The variance explained by trial number increased ($p=0.01$). In sum, AC inactivation selectively affected the early auditory-related component of sound-evoked V1 activity, but not the motor-related component.

Behavioral dominance of audition over vision

How do auditory and motor signals impact concurrent visual processing? In a subset of sessions in MST mice, we interleaved trials in which both visual and auditory stimuli changed simultaneously with standard trials presenting a unisensory change and catch trials. In these mice, visual and auditory feature changes were associated with different motor actions. The modalities therefore acted as competing inputs and a simultaneous change presented the animal with a conflicting situation. Therefore we first describe the behavior.

In behavioral sessions we presented four levels of visual and auditory feature change that matched in subjective saliency across modalities (Meijer et al., 2018; Song et al., 2017). Stimuli were taken as the x-axis positions corresponding to the same positions along the psychometric function of each modality in unimodal trials and thus matched performance (Fig. 4.5a; subthreshold (*sub*), threshold (*thr*), suprathreshold (*sup*), maximal (*max*)). Saliency-matched conditions led to comparable increases in pupil dilation (Ext. Data Fig. 4.9a,b). In conflict trials, animals predominantly chose the auditory lick spout (Ext. Data Fig. 4.9c). Behavioral choice scaled with the saliency of the sensory input, but in saliency-matched conditions auditory choices dominated (Fig. 4.5b,c; Ext. Data Fig. 4.9d). A saliency-matched dominance index (smDI) was computed as a ratio of auditory to visual choices for saliency-matched conflict conditions (Fig. 4.5d). The smDI

was significantly higher than zero (indicating auditory dominance), confirming auditory dominance (smDI: 0.33, Wilcoxon signed rank test, p=0.0014, n=17 mice, based on 97 sessions, 51.932 trials). Audition thus dominates behaviorally over vision in our modality identification task and this is partly explained by faster auditory processing (Ext. Data Fig 4.9).

Dissociation of auditory and visual processing: preserved orientation coding during conflicting multisensory inputs

To examine how auditory- and motor-related inputs intersect with concurrent visual processing, we sampled neural activity during conflict trials of threshold and maximal audiovisual saliency in 65 out of 122 recording sessions with MST mice. The time window of analysis was broadened to 0-500 ms to investigate continued interaction between visual, auditory and motor-related processing.

We first compared trial-averaged activity during multisensory conflicts with saliency-matched unimodal trial types (Fig. 4.5e). Neurons that responded during visual or auditory trials continued doing so during conflict trials. In the regression model, visual, auditory and motor-related predictors continued to predict V1 firing rate in held-out conflict trials (Fig. 4.5f).

These results are in line with our finding that visual, auditory and motor components evoked activity in largely distinct neuronal subsets of V1 (Fig. 4.3f). Briefly revisiting the unisensory trials, population activity would be expected to distinguish between visual and auditory trials. Indeed a decoder trained to

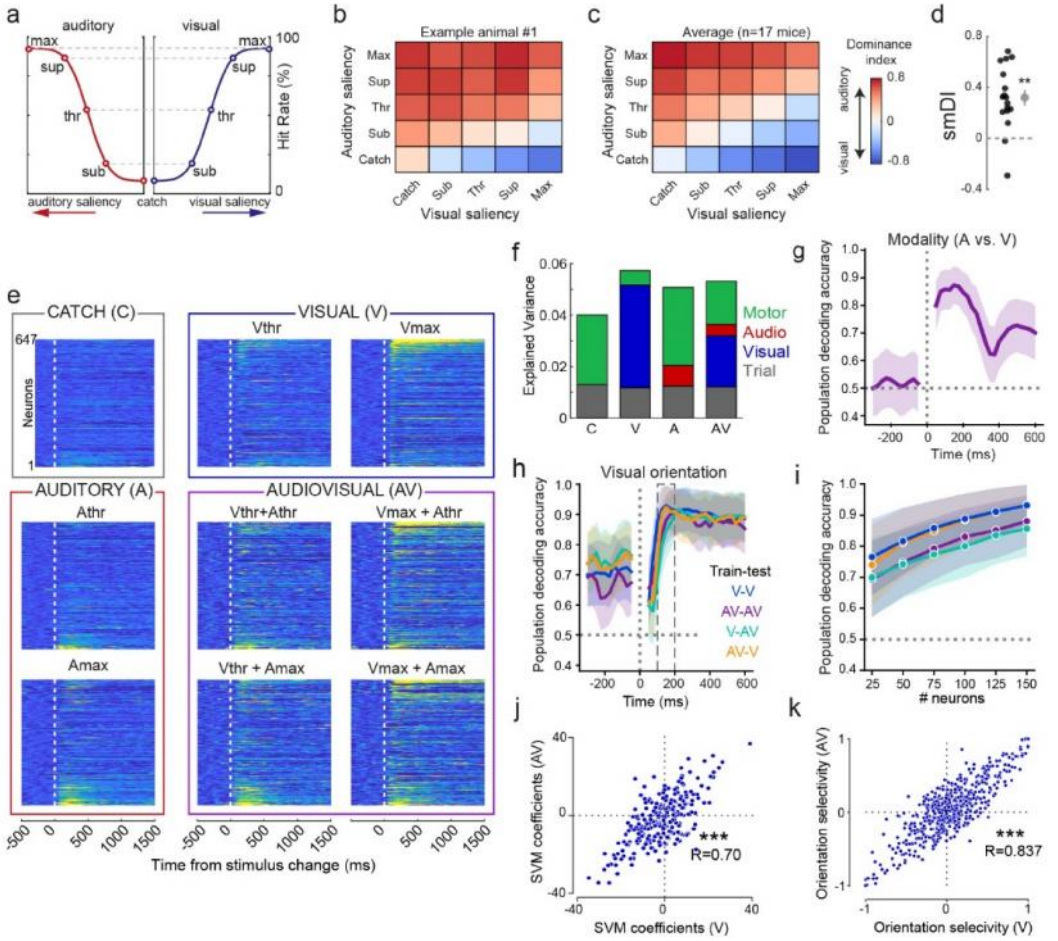


Figure 4.5: Preserved orientation coding in V1 during multisensory trials showing auditory behavioral dominance. a) To equalize subjective saliency during conflict trials, four levels of auditory and visual change were taken at equal positions along this schematic psychometric curve. All data in this figure are from MST animals. **b)** Dominance index (DI) as a heatmap across unimodal and conflict trials for an example animal showing auditory dominance. A positive (red) DI indicates a higher fraction of auditory choices relative to visual. **c)** Heatmap as in (b), but averaged across all animals (n=17). **d)** Saliency-matched Dominance Index (smDI) taken by averaging the DI of all saliency-matched conflict conditions (conditions along top-right to bottom-left diagonal in heatmaps of (b) and (c), excluding catch trials). Each dot is one mouse. Grey dot and error bar are mean ± SEM. **p<0.01. **e)** Heatmap of z-scored firing rate across V1 neurons (each row in each heatmap is a neuron). Each heatmap shows a trial type, with columns increasing in auditory saliency (no change, threshold, maximal) and rows increasing in visual saliency. Lower right 4 panels indicate conflict trials. Neurons are sorted by response magnitude difference between unimodal visual and auditory trials. During audiovisual trials, subsets of visually responsive (upper rows in each heatmap) and auditory responsive neurons (bottom rows) respond similarly as during unimodal trials. **f)** Total explained variance in firing rate of V1 cells for each trial type, with each color indicating the contribution of each predictor subset. Result obtained from the regression model. C: catch trials. **g)** Decoding performance over time for decoders trained to discriminate the modality of unimodal trials (visual versus auditory) from V1 pseudopopulation activity (n=150 neurons). Line and shading

indicate mean and 95% CI. Grey dotted line is chance level. **h)** Same as (g), but for performance of decoders trained to discriminate visual orientation. Decoders were trained and tested on held-out test trials of the same trial type (within condition: V-V and AV-AV) or trained on visual and tested on held-out audiovisual trials, or vice versa (cross-condition: V-AV and AV-V). Baseline performance results from a subset of neurons persistently coding orientation, see Fig. 2h,m. Line and shading indicate mean and 95% CI. **i)** Quantification of orientation decoding performance averaged over dashed time window in (h) for different pseudopopulation sizes. **j)** Orientation decoding weights (SVM coefficients) at t=0.175 sec from decoders trained separately on visual or audiovisual trials. Each dot is a neuron. **k)** Orientation selectivity during visual (Vmax) and audiovisual trials (Vmax+Athr or Vmax+Amax).

discriminate the modality of unimodal trials from V1 pseudopopulation spiking data (pseudopopulation size n=150 neurons) did so with high accuracy (Fig. 4.5g). Returning to conflict trials, the auditory stimulus and orofacial movements explained more variability in conflict trials than visual stimuli did (Fig. 4.5f). With these multimodal inputs competing, we wondered how visual representations would be affected. We trained the same population decoder to discriminate visual orientation in unimodal visual and multimodal (conflict) trials and tested performance on held-out test data of the same or different trial type (i.e. with or without auditory stimuli). For all conditions tested, orientation decoding performance increased after visual stimulus change (Fig. 4.5h) and was comparable between visual-only and conflict trials, also when tested for various population sizes (Fig. 4.5i). The contributions of individual neurons (SVM coefficients) were strongly correlated between visual and audiovisual trials (Fig. 4.5j; $R=0.70$, $F(1,247)=236.24$, $p=7.28*10^{-38}$). This population-level finding matched with single-neuron analyses, where orientation selectivity during visual and conflict trials was strongly correlated in a similar, but even stronger manner (Fig. 4.5k; $R=0.837$; $F(1,529)=282.61$, $p=3.97*10^{-51}$). Therefore, while sounds dominate behavioral choice and V1 variance, visual feature coding is only minimally affected.

Discussion

Several studies have reported sound-evoked activity in V1, but often without controlling carefully and systematically for sound-evoked behavioral changes. We found that a large part of sound-evoked activity in V1 appeared correlated to orofacial movements, in line with a recent report (Bimbard et al., 2021). However, next to motor-related activity, distinct early auditory-related activity was observed in V1. These inputs likely reflected auditory sensory-evoked inputs as they had short-onset latencies in spiking and LFP data, did not correlate to orofacial movements, and were reduced after AC inactivation (Fig. 4.2d,e; 4.3b,f,g; 4.4f,h; Ext. Data Fig. 4.7g,h). Rather than there being a single external input of a purely sensory or purely behavioral origin, both auditory and motor-related inputs reach primary visual cortex, and these are segregated in time and space (via different V1 subsets). Sounds can thus lead to strong activity changes in visual cortex through multiple pathways. In addition to the dissociability of auditory and motor

processing, we show that auditory and visual streams remain largely segregated, most poignantly illustrated by the preserved orientation coding during multisensory conflict trials in which auditory inputs dominated behaviorally.

Early auditory-related activity in visual cortex

Even when motor-related influences were accounted for, short-latency AC-dependent signals were present in V1. Onset latencies of sound-evoked activity were comparable to the prior literature with 18 ms in AC (11 ms in Sakata and Harris, 2009) and 27 ms in V1 (35.8 ms in Iurilli et al., 2012). This latency difference is compatible with mono- or disynaptic AC-V1 connections. The specific reduction of auditory-related activity in V1 after bilateral pharmacological inactivation of AC further establishes the efficacy of the AC-V1 pathway during task performance (Fig. 4.4). Muscimol injections also increased the variance explained by trial number; it may be speculated that this effect was caused by slow fluctuations in excitability of thalamocortical networks affecting V1 activity across the session, even though the mean firing rate was comparable to control recordings (Fig. 4.4c). AC inactivation impaired but not abolished the detection of frequency changes, in line with a modest role for AC in pitch discrimination (Ceballo et al., 2019). Given that primary auditory and visual cortices are monosynaptically connected and sound-evoked activity in V1 has been reported under anesthetized conditions (Henschke et al., 2015; Ibrahim et al., 2016; Iurilli et al., 2012), our findings underscore the efficacy of direct auditory inputs to visual cortex in awake behaving mice, which is apparently not overshadowed by motor and arousal effects or task-dependent factors that could in principle regulate the strength of these inputs (cf. Knopfel et al. 2019).

Our anatomical tracing and LFP results are in line with previous reports that sounds can modulate visual cortex through both superficial L1 projections (Ibrahim et al., 2016) and deeper inputs (Iurilli et al., 2012). These inputs may modulate cortical activity through translaminar dendrites (e.g. targeting L1 apical dendrites of L2/3 and L5 neurons) as well as through translaminar inhibitory circuits, modulating activity in supragranular and infragranular layers (Fig. 4.3) (Ibrahim et al., 2016; Iurilli et al., 2012; Knöpfel et al., 2019; Meijer et al., 2017).

Visually evoked activity in auditory cortex, on the other hand, was strikingly absent. This asymmetry – or predominance of auditory crossmodal influences on visual cortex relative to visual-to-auditory cortex influences - matches earlier physiologial and anatomic studies (Budinger and Scheich, 2009; Campi et al., 2010; Ibrahim et al., 2016; Iurilli et al., 2012; Oh et al., 2014), but is noteworthy because it now is shown to hold in a task setting where auditory and visual changes were of equal behavioral relevance. This asymmetry, however, might also depend on the stimulus characteristics (Chou et al., 2020).

Behavioral component of sound-evoked activity in visual cortex

A large part of sound-evoked activity in V1 correlated to orofacial movements. Already 60-100 ms after auditory changes, motor activity started (Fig. 2a) which

correlated with V1 spiking activity (Fig. 4.2e; 4.3f; 4.5f) and LFP (Ext. Data Fig. 4.3g). These orofacial movements were themselves stimulus-specific, such that auditory stimulus identity could be decoded from video footage (Fig. 4.2j-l). Although the auditory stimuli were a weighted combination of different tones, 'motor-tuning' might arise through differences in subjective saliency, arousal, or aversion to particular auditory frequencies. These results, including the frequency-tuned responses in V1, are in agreement with a recent study showing that different sound clips were associated with stereotypical orofacial movements across mice, which correlated to V1 activity (Bimbard et al., 2021). These and the present results strongly argue for cautious interpretations of multisensory interactions in awake subjects and underline the need to carefully monitor behavioral state.

We labeled neural activity changes related to orofacial movements as motor-related. It is unclear whether these are better interpreted as corollary discharge signals to V1 predicting (or otherwise relating to) visual consequences of motor movements (Guitchounts et al., 2020; Leinweber et al., 2017; Pennartz et al., 2019; Schneider et al., 2014), or as internal state changes associated with arousal levels, which may be linked with increased movement (Niell and Stryker, 2010; Vinck et al., 2015c). Visual and auditory stimuli evoked approximately similar levels of pupil dilation (Ext. Data Fig. 4.9a), but auditory stimuli could in addition elicit fast arousal responses, known to originate from intralaminar thalamic nuclei (Minamimoto and Kimura, 2002; Van der Werf et al., 2002). As corollary discharge signals might be temporally shifted relative to video-observed movements (Leinweber et al., 2017), a methodological limitation is that our regression model does not take into account nonlinear or temporally shifted relationships between orofacial movements and V1 activity. Whether sounds predominantly activate arousal or sensorimotor signaling in visual cortex is a question for future investigation.

Preserved orientation coding during auditory behavioral dominance

Mice reported auditory over visual stimulus changes during saliency-matched conflict trials in a similar manner as in Song *et al.* (2017) (but see (Coen et al., 2021) for balanced audiovisual weighting). During these conflict trials, the combination of auditory and motor-related components explained more variability in firing rate than visual stimuli, even though orientation coding was preserved. In other words, both auditory and motor influences on V1 activity appear to be largely orthogonal to visual feature representations, cf. (Montijn et al., 2016; Stringer et al., 2019). This preserved orientation coding matches well with our observation that visual, auditory, and motor-related activity occurs in rather segregated cell populations (Fig. 4.3f) and underscores their dissociability. Primary visual cortex thus supports (relatively independent) parallel encoding of signals related to different sensory and motor modalities. This segregation somewhat contrasts with studies reporting predominantly jointly responsive neurons (Bizley et al., 2007; Knöpfel et al., 2019), which may relate to our behavioral task in which auditory and visual cues were explicitly not to be integrated.

Beyond cue integration

Our work differs in an important way from the large body of literature on multisensory cue integration (Fetsch et al., 2013; Meijer et al., 2019; Stein and Stanford, 2008), where for instance crossmodal inputs are interpreted to improve the inference of grating orientation (Ibrahim et al., 2016; Nikbakht et al., 2018; Williams et al., 2021). In the MST task, the two modalities were not jointly informative about the same external variable (e.g. they were not jointly indicating a source location, heading direction or stimulus rate), but rather auditory signals were statistically uninformative about visual features (and vice versa), in contrast to e.g. (Garner and Keller, 2022; Lippert et al., 2007; Meijer et al., 2018, 2020; Sheppard et al., 2013). Instead, animals needed to monitor potential changes in both modalities across time, without a trial onset cue, and discriminate in which modality a change occurred.

Our finding that auditory changes not only modulated, but evoked spiking in V1 of naive and visually trained mice as well as in mice trained to discriminate auditory and visual signals, suggests a broader role for auditory signals in V1 than only sharpening visual tuning. As shown elsewhere, auditory signals in V1 do not become causally important for audition during MST training, as V1 optogenetic inhibition impacts visual but not auditory change detection performance (Oude Lohuis et al., 2022). Even if auditory inputs to visual cortex are not directly relevant for detecting single visual features, they are hypothesized to fit in a broader view of sensory cortical function (Meijer et al., 2019; Pennartz, 2015, 2009; Petro et al., 2017), where crossmodal interactions serve to orchestrate perception across a larger cortical network, guide crossmodal attention, and inform visual processing in distributed networks about ongoing auditory events. In this respect, it is interesting that the auditory component was stable across cohorts (Ext. Data Fig. 4.5c), consistent with a basal, rather than task-specific function. Part of this orchestration may reside in predictions that fast auditory processing conveys upon vision; another part may relate to the observation that perceptual phenomenology is qualitatively rich, and thus requires segregation as well as integration of sensory modalities (Pennartz, 2015, 2009).

In sum, to correctly interpret (multi)sensory-evoked activity, careful dissociation of sensory and motor origins is necessary. Through this dissociation it becomes clear that sound evokes inputs from auditory to visual cortex that are fast and transient, as well as to later, secondary motor modulations through sound-evoked body movements. The associated activity patterns temporally overlap somewhat and co-exist with visual processing, although this remains dissociable as apparent from population coding of visual grating orientation. An exciting direction of future research will be to understand how these multiple signals co-exist to contextualize sensory input to generate meaningful information processing and behavior.

Acknowledgments

We thank Jorrit Montijn and Guido Meijer for useful comments on the manuscript, D. Sridharan for providing code for the multi-alternative detection model; C. Rossant, members of the Cortex Lab (UCL) and contributors for Klusta and Phy spike sorting software; Andriana Mantzafou, Klara Gawor, and Alexis Cervàn Canton for assistance in behavioral training. This work was supported by the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement 945539 (Human Brain Project SGA3) to C.M.A.P. and by the FLAG-ERA JTC 2019 project DOMINO (co-financed by NWO) to U.O.

Author contributions
Conceptualization, M.O.L., U.O., C.M.A.P.; Methodology, Formal Analysis, Visualization, M.O.L.; Additional analysis, P.M.; Writing, M.O.L., U.O., C.M.A.P.; Supervision & Funding, U.O., C.M.A.P.

Declaration of interests
The authors declare no competing interests.

Data availability
All behavioral and neural data related to this study are available on FigShare (FINAL LINK TO BE INSERTED).

Code availability
All code related to this study is publicly available on GitHub (FINAL LINK TO BE INSERTED).

Methods

Animals
All animal experiments were approved by the Dutch Commission for Animal Experiments and by the Animal Welfare Body of the University of Amsterdam. Thirty-three male mice were used from different genotypes: wildtype C57BL/6, PVcre (JAX 008069), and PVcre/TdTomato (JAX 027395). Mice were at least 8 weeks of age at the start of experiments. Mice were group-housed under a reversed day-night schedule and all experimental procedures were performed during the dark period (8:00 – 20:00).

Head bar implantation
Before the start of any experiment, a custom-made titanium head-bar was implanted to allow head fixation. Mice were anesthetized with isoflurane and fixed in a stereotaxic apparatus. A circular patch of skin was removed to expose and disinfect the skull. A circular head bar (inner diameter 10 mm) was positioned over the skull to include bilateral V1 and AC and glued and cemented to the exposed

skull. After a recovery period of 2-7 days, mice were habituated to handling and head-fixation before start of the training procedure.

Behavorial tasks

Mice were water-deprived throughout the course of experiments and earned their daily ration of liquid by performing the behavioral task. In the case of low performance, daily intake was supplemented to a minimum of 0.025 ml per gram of body weight. During a session, mice were headfixed their bodies were positioned in a cylindrical holder. Two lick spouts were positioned symmetrically on the left and right side within reach of their tongue. Licks were detected by capacitance-based (during training) or piezo-electric based detectors (during recordings). Correct licks were immediately rewarded with 5-8 µl of liquid reward (infant formula; Nutrilon) delivered through the same lick spout using gravitational force and solenoid pinch valves (Biochem Fluidics, Boonton, USA).

Audiovisual change detection task

In the audiovisual change detection task, auditory and visual stimuli were continuously, without pre-cueing, presented throughout a behavioral session. During visual trials a feature (the orientation) changed, after which this new feature (the post-change orientation) continued to be shown. Similarly, auditory trials consisted of a change of one auditory stimulus to another.

Stimuli

Visual stimuli were drifting square-wave gratings with a temporal frequency of 1.5 Hz and spatial frequency of 0.08 cpd at 70% contrast (35 cd/m² luminance difference between bright and dark). In trials with a visual change the orientation of the drifting grating was instantaneously changed (e.g. from 60˚ to 90˚) while preserving its phase. Visual stimuli were presented with a 60 Hz refresh rate on an 18.5-inch monitor positioned at a straight angle with the body axis from the mouse at 21 cm from the eyes.

Each auditory stimulus was a weighted combination of five pure tones at harmonic frequencies: a center tone, as well as two lower and two higher harmonics (octaves below and above the center tone). If f_0 is the center tone, then: $f_{-2} = 1/4 * f_0$, $f_{-1} = 1/2 * f_0$, $f_0 = f_0$; $f_{+1} = 2 * f_0$; $f_{+2} = 4 * f_0$. We name each auditory stimulus after the frequency of its center tone. All frequencies were expressed in scientific pitch as powers of 2 with the center tones spanning from 2^{13} Hz (=8372 Hz) to 2^{14} Hz (=16744 Hz). An example stimulus, $2^{13.5}$ (named by center tone), was therefore composed of five pure tones of $2^{11.5}$, $2^{12.5}$, $2^{13.5}$, $2^{14.5}$, and $2^{15.5}$ Hz. The weight with which each tone was present was taken from a Gaussian distribution across all tones for all stimuli, centered at $2^{13.5}$ (=11585 Hz). Lower and higher harmonics thus contributed less to the auditory stimulus than the center tone. Because of this fixed weight distribution, stimuli with higher center tone frequency have decreasing weights for higher harmonics and increasing weights for lower harmonics. Stimuli with higher center frequency are thus increasingly made up of lower frequency components to the point of arriving at the starting stimulus (see also Ext. Data Fig. 4.1). This

auditory stimulus design with harmonics and fixed weights was inspired by the Shepard tone illusion (Shepard, 1964). However, in contrast to this illusion, our stimuli were static and not sweeping across frequencies, and the original illusory aspect of a tone ever-increasing (or decreasing) in pitch was not exploited. The primary reason for this auditory stimulus design was the circular nature of the stimulus set, which mirrored the visual stimulus set with drifting gratings in all orientations.

During trials with an auditory change, one stimulus was changed instantaneously to another. This resulted in a shift in spectral power to five new frequencies which appeared to the mouse as an increase or decrease in pitch. Auditory changes were expressed as partial octaves, with ½ octave maximally salient and the minimal change used was 1/256 partial octave. The degree of frequency/octave change determined the auditory saliency and was varied across experimental conditions. During auditory stimulus changes, the phase across all tones was preserved. Stimuli were presented with a sampling rate of 192 kHz. Stimuli were high-pass filtered (Beyma F100, Crossover Frequency 5-7 kHz; Beyma, Valencia, Spain) and delivered through two bullet tweeters (300 Watt) directly below the screen. Note that this high-pass filter eliminated the lowest frequency components of the Shepard stimuli, and left the mid and high frequency components intact (those that span the sensitive part of the mouse hearing range, 8-16 kHz). This was done to prevent damage to the specialized tweeters that we used, but did not affect the animals' ability to report even very small differences between subsequently presented Shepard tones. Sound pressure level was calibrated at the position of the mouse and volume was adjusted per mouse to the minimum volume that maximized performance (average ±70 dB).

In an earlier cohort of mice (N=13/33) and for the audiovisual detection task (n=3, see below), the Shepard tones (1) were expressed in absolute Hz (e.g. an auditory trial with Δ2kHz changed from 8 kHz to 10 kHz), (2) had 9 instead of 5 harmonics, (3) were presented with a sampling rate of 48 kHz and (4) were not phase-preserved during a change in auditory frequency. We observed no qualitative or quantitative differences in both neural and behavioral results between the cohorts and the data was pooled for all analyses.

Trial types

Trials simply consisted of an instantaneous feature change (visual, auditory or audiovisual) and an ensuing reward window. Trial onset was defined by an instantaneous change in the visual or auditory stimulus, or no change (catch trial). All analyses are relative to this stimulus change. Trials were separated by an inter-trial interval randomly taken from an exponential distribution (mean 6, minimum 3, and maximum 20 seconds). Trial types were pseudorandomly ordered by block-shuffling per 10 trials (8% catch trials=no change, 46% visual trials, 46% auditory trials). In sessions with multimodal conflict trials these replaced unimodal trials (see below).

Task versions

Animals were assigned to one of three versions of the audiovisual change detection task in which the visual and auditory stimuli were identical and only the reward contingencies varied (i.e. which stimuli were rewarded). This led to a controlled manipulation of the behavioral relevance of the stimuli as well as differences in the amount of instructed movements (goal-directed licking). Noncontingently exposed (NE, n=7) animals were not rewarded for licking after auditory or visual stimulus changes, but obtained rewards for licks during hidden 'response windows' that were temporally offset from the stimuli. This resulted in spontaneous licking behavior at the two spouts that was occasionally rewarded. Unisensory-trained (UST, n=4) animals were trained to only report visual changes and ignore auditory changes. Auditory stimuli and changes were presented throughout all sessions, but were not associated with reward and changes were temporally decorrelated from the task-relevant visual trials (no accidental conflict trials were programmed). Multisensory-trained animals (MST, n=17) animals were trained to detect and identify one of both modalities. Animals were required to respond in a lateralized manner to each modality: lick to one side to report visual changes, to the other side for auditory changes (modality-side pairing was counterbalanced across mice). In other words, mice were required to simultaneously monitor both the auditory and visual modality and identify the sensory modality in which a change occurred. As we performed additional experiments with animals from the MST cohort, this resulted in a higher number of animals in the MST cohort.

Psychometric performance

Animals in the NE cohort were accustomed to spontaneous licking behavior irrespective of sensory stimuli in a few sessions. Animals in the UST and MST cohorts were trained over the course of several weeks in which progressively more difficult trial types (lower saliency) were introduced and reward size was lowered until performance stabilized. To match the subjective salience of auditory and visual stimuli across mice and modalities we chose intensities according to their unimodal hit rates (Meijer et al., 2018; Song et al., 2017). For each trained animal we established perceptual sensitivity by presenting five levels of auditory and visual saliency (amount of change) that spanned the perceptual range for three consecutive behavioral sessions. We fit the concatenated data of these three sessions with a cumulative normal distribution per modality with four free parameters (Meijer et al., 2018):

$$f(x) = \gamma + (1 - \gamma - \lambda)\left(\frac{1}{2}\left[1 + e^{\frac{x - \mu}{\sigma\sqrt{2}}}\right]\right) \qquad (E\ .4.1)$$

Here, γ describes the false alarm rate (spontaneous licks during catch trials), λ the lapse rate (misses at maximal saliency), μ the mean of the cumulative normal distribution (perceptual threshold), and σ the standard deviation (sensitivity to

variations of stimulus intensity). Having established the psychometric function per mouse, we took four levels of saliency per modality at fixed points along the psychometric function: subthreshold (μ-σ; *sub*), threshold (μ; *thr*), suprathreshold (μ+ σ; *sup*), and maximal saliency (*max*).

Conflict trials

In a subset of sessions we introduced multimodal trials in which auditory and visual stimuli simultaneously changed (conflict trials, making up 25% of all trials; replacing a fraction of unimodal trials). These multimodal trials were introduced in experiments with NE and MST mice. Multimodal trials were omitted in experiments with UST mice as systematic pairing of auditory and visual changes would render the auditory stimuli predictive of concurrent visual changes, while in these mice we aimed to study the processing of auditory stimuli under behaviorally irrelevant conditions.

For NE mice, multimodal trials were presented during nearly all recording sessions (n=27/28 sessions). For MST mice, we first quantified behavioral choice during conflict trials in a set of sessions without recordings for all combinations of visual and auditory saliencies (4x4=16 saliency combinations). This protocol was repeated for 4-7 sessions per animal and the data was averaged for analysis (total N=97 sessions; N=17 mice). For MST mice, the side of the first lick was registered as the animal's choice. To maintain consistency in task rules (i.e., that a change predicts reward) licking to both spouts was rewarded. In a separate set of experiments (N=20 sessions, 4 mice) we systematically varied the stimulus-onset-asynchrony (SOA) of the auditory and visual change in conflict trials. In these trials we presented only one combination of auditory and visual subjective intensity (both at *threshold* saliency) and used the temporal offsets: -300 ms, -100 ms, -30 ms, 0, +30 ms, +100 ms, +300 ms (where negative values mean that the auditory stimulus changes first). Multimodal trials were also introduced in a subset of the recording sessions (N=65/122 recording sessions). During recording sessions only two levels of saliency were used.

Due to our continuous stimulus design, we were constrained in the timing of visual and auditory changes. A constraint on the auditory change resulted from the fact that the auditory stimulus was changed only when all component tones were aligned in phase. This was done to avoid inducing artefacts changing the frequency of pure tones out of phase. For the visual domain, a constraint on precise timing resulted from the refresh rate of the monitor (60 Hz). To achieve maximal alignment in audiovisual trials, we first computed the future timestamp of the phase-preserved change in auditory frequency. The visual stimulus changed at the frame closest to that timestamp. Stimulus-onset asynchrony was therefore maximally 8.33 ms; half the duration of the interframe interval (0.5 * 16.67 ms). The direction of this misalignment varied and was small relative to the timescale of analysis of conflict trials (0-500 ms).

Stimuli during recording sessions

During recording sessions, the trial type conditions were limited to get sufficient repetitions. First we limited trials to use only two levels of saliency, *threshold* and *maximum*. Threshold intensity was obtained through psychophysical experiments per modality and per mouse (described earlier) and maximum intensity was always fixed (90 deg and ½ octave). Second, we fixed the auditory and visual stimulus identities, such that changes occurred only between 4 visual stimulus orientations and 4 auditory frequencies (A, B, C, D). The distance between A and B and between C and D was at threshold level, while the distance between A and C and between B and D was maximal. An example stimulus set was 90, 97, 180, 187 (in degrees), and 2^{13}, $2^{13.03125}$, $2^{13.5}$, $2^{13.53125}$ (in partial octaves; Ext. Data Fig. 4.1d,e). Auditory and visual stimuli therefore jumped back and forth between four orientations and frequencies across trials, providing reliable estimates of tuning to specific features. For naive mice we used threshold values that matched those from trained animals. Nearby stimuli (A and B, as well as C and D) evoked highly similar activity patterns and for all analyses, the two nearby stimuli were grouped (i.e. set A/B and set C/D).

Engaged versus passive

In a subset of MST mice and in a separate set of sessions (N=5), we combined task-engaged and passive blocks within the same recording session. During active blocks the lick spouts were accessible, while in passive blocks the lick spouts were manually positioned out of reach. The stimuli as well as the temporal statistics and trial type distributions were the same for the active and passive blocks within that recording session. In some sessions we implemented one passive and one active block. In other sessions multiple active and passive blocks were alternated (n=5-7 blocks). The order of passive and active blocks was counterbalanced across sessions. No differences were found between single versus multiple alternating blocks and data were pooled. Only video data was analyzed (Ext. Data Fig. 4.2c-e).

Audiovisual stimulus detection task

We trained additional mice (n=3) on a simpler variant of the change detection task, in which we could present a larger set of stimuli. In this detection task animals had to detect stimulus presence, rather than stimulus change. Each trial consisted of a blank intertrial interval (gray screen, no sound) drawn randomly from the same distribution as the change detection task and a stimulus window (1.5 seconds) during which a reward could be obtained for licking the correct lick spout. Analogous to the change detection task, animals had to report presence of a stimulus and its modality by directed licks to either the left lick spout (visual) or right (auditory). Modality-side pairing was the same for the three animals. Four trial types were used: visual (41% of trials), auditory (41%), catch (no stimulus, 8%), and conflict trials (10%). Conflict trials were not analyzed. The same stimulus set (full-field drifting gratings and Shepard tones) was used as in the change detection task. The saliency of each trial was now determined by grating contrast, while auditory saliency was determined by sound volume.

Similar to the change detection task, we first established psychophysical performance of each mouse in a series of behavioral sessions with variable visual and auditory saliency across the perceptual range (Meijer et al., 2018). Subsequently, for recording sessions a fixed saliency was chosen at threshold level in the same way as for the change detection task, resulting in substantial numbers of hits and misses. Eight orientations (spaced 45˚) and eight frequencies (center tone spacing at 1 kHz) were presented at this saliency level. Visual contrast levels for the three animals were 9.7%, 10%, 13%. Auditory saliency levels were 56, 78, 82 dB. Note the higher volume used for two animals. This cohort of animals was of old age at the time of experiments (40-43 weeks) and two of these animals showed a progressive decline in their sensitivity to auditory stimuli of lower and intermediate volumes. This is in line with age-related hearing loss reported in C57BL/6 mice (Henry and Lepkowski, 1978; Spongr et al., 1997). We included these mice irrespective of their decreased hearing sensitivity as we titrated auditory intensity to result in similar subjective saliency (hit rates were comparable for both visual and auditory trials: 49% and 46%, respectively) and we focused in this task on correlated feature tuning of V1 neurons, not on performance aspects.

Electrophysiological recordings

On the day before the start of extracellular recording sessions, mice were anesthetized with isoflurane and small (~200μm) craniotomies were made using a dental drill over the areas of interest. Areas of interest were binocular V1 (relative to lambda: AP 0.0, ML ± 3.00 mm) and AC (relative to bregma: AP -2.6 mm, ML ± 4.3 mm). Craniotomies and recordings in medial prefrontal cortex and posterior parietal cortex were also performed, but data from these areas were not analyzed here. Extracellular recordings were performed on consecutive days with a maximum of 4 days per mouse to minimize damage to the circuitry. Each recording session, up to three microelectrode arrays (silicon probes) of 32 or 64 channels (NeuroNexus, Ann Arbor, USA – A1x32-Poly2-10mm-50s-177, A4x8-5mm-100-200-177, A1x64-Poly2-6mm-23s-160) were slowly inserted into their target area. We approached V1 perpendicularly to the cortical surface and lowered the silicon probe until all recording sites spanned the cortical layers of V1. Because of the circular headbar (inner diameter 10 mm) we used, craniotomies were located slightly medially on the skull surface and AC was approached with an angle approximately 30˚ away from the midline and with 64-channel laminar probes that spanned 1450 μm. Due to the span and angle of approach, we recorded multiple subfields of the auditory cortex. In a subset of animals, on the last day of recordings the probe was covered in DiI (ThermoFisher Scientific) to facilitate post hoc reconstruction of the electrode tract. Neurophysiological signals were pre-amplified, bandpass filtered (0.1 Hz to 9 kHz), and acquired continuously at 32 kHz with a Digital Lynx 128 channel system (Neuralynx, Bozeman, MT). The start of the behavioral task commenced at least 15 minutes after probe insertion to allow for tissue stabilization.

Video monitoring

A near-infrared monochrome camera (CV-A50 IR, JAI, Copenhagen, Denmark) was coupled to a zoom lens (50 mm F/2.8 2/3" 10MP, Navitar, Rochester, USA) and positioned at approximately 30 centimeters from the mouse to capture the lick spouts and face of the mouse within a single view. The left side of the face was illuminated with an off-axis infrared light source (IR-LEDs, 850 nm) positioned to yield high contrast illumination of both the eye and whisker pad. A frame grabber acquired images of 752x582 pixels at 25 frames per second. With this acquisition rate the timing precision for the facial motion and pupil size and location was about 40 ms.

Optogenetics

Mice (N=5 NE animals, 20 weeks old) were subcutaneously injected with the analgesic buprenorphine (0.025 mg/kg) twenty minutes prior to surgery. During surgery, mice were maintained under isoflurane anesthesia (induction at 3%, maintenance at 1.5–2%). We aimed at infecting bilateral AC and centered our injection at primary auditory cortex (A1). We performed a small craniotomy over bilateral primary auditory cortex (A1 relative to bregma: AP -2.60 mm, ML ± 4.30 mm) using an ultra-fine dental drill and inserted a glass pipette backfilled with AAV2-CamkIIa-hChR2(H134R)-EYFP (titer: 3×10^{12} viral genomes per ml, 26969-AAV2, Addgene). AC was approached similar to extracellular recordings. Four injections of 13.8 nl were made using a Nanoject pressure injection system (Drummond Scientific Company, USA) at two depths: two at 1200 µm and two at 1000 µm below the dura (ending up in A1 because of the angle). Each injection was spaced apart from the next one by 5 minutes to promote diffusion and prevent backflow. After viral injections, the recording chamber was covered with silicon elastomer (Picodent Twinsil) and mice were allowed to recover.

After 4-6 weeks to allow for viral expression, the silicon elastomer was removed during recording sessions. To photostimulate AC, a fiber-optic cannula (inner diameter 200 um, numerical aperture 0.48, DORIC Lenses, Quebec, Canada) was positioned directly over AC. The fiber-optic cannula was sealed with black tape, leaving only the tip exposed to prevent light from reaching the eye of the mouse. A fiber optic patch cord connected the cannula to a 473 nm laser (DPSS 473nm H300, Eksma Optics, Vilnius, Lithuania). A shutter (LS6 Uniblitz, Vincent Associates, Rochester, USA) controlled light delivery and was located in a sound-insulated box distal from the experimental setup to prevent any auditory-evoked activity. AC was stimulated with 10 ms pulses with variable laser power (0-20 mW total power) and variable frequencies (5, 10, 20 and 50 Hz).

Muscimol inactivation of AC

On the day before the start of muscimol experiments a craniotomy over AC was made using the same coordinates as for the recordings and optogenetics. To inactivate AC, 300 nl of muscimol solution (10 mM in saline, pH 7.2; Sigma Aldrich) or saline solution (control) was injected in bilateral AC. AC was approached with the same coordinates (centered at primary auditory cortex) and

angle of approach as with extracellular recordings or viral injections. Glass micropipettes were backfilled and slowly inserted through the craniotomy. Three injections of 50 nl were done at 1200 µm and three at 1000 µm below the dura using the Nanoject injection system, with one minute spacing between each injection. AC inactivation was verified using multi-unit recordings in A1, taking V1 as a control area. In a subset of animals we injected BODIPY TMR-X conjugated muscimol (ThermoFisher Scientific; Catalog number: M23400) during the last session to assess the localization and spread of AC muscimol injections. Injecting 300 nanoliter led to localized expression in AC, primarily in A1 (Fig. 4.4a), whereas we found that injecting a larger volume of 500 nanoliter led to reduction of spontaneous activity in V1 after approximately 10 minutes, which we interpreted as extended diffusion. During experiments with simultaneous recordings the silicon probe was first inserted to stabilize, then we performed muscimol injections, and directly after these the behavioral experiment was started. Muscimol or saline injections were performed on alternating days. We observed comparable behavioral performance on days after muscimol experiments.

Histology

At the end of the experiment, mice were overdosed with pentobarbital (>100 mg/kg) and perfused (4% paraformaldehyde in phosphate-buffered saline, pH 7.2). The brains were recovered for histology to verify viral expression and recording sites. Coronal sections were cut at 50 µm and overlaid with the matching reference section from the atlas (Paxinos and Franklin, 2004). Flattened cortical sections were cut at 50 µm, prepared as described previously (Lauer et al., 2018), and V1 and AC were identified based on cell densities aligned to reference maps (Gămănuţ et al., 2018).

DATA ANALYSIS

Unless otherwise stated, all data were analyzed using custom-made software written in MATLAB (The MathWorks, Natick, USA) or Python (analysis on population decoding only). All code and data will be made available on Github (insert link) and FigShare. Given that sound-evoked activity consisted of both auditory and motor-related activity in V1 of mice from all three cohorts, data from all cohorts were combined, unless otherwise specified.

Video analysis

To capture and describe orofacial movements from video recordings, the principal components of motion across the video frames were extracted using FaceMap (Stringer et al., 2019). Briefly, the video was spatially downsampled (1 every 4 pixels) and singular value decomposition was applied on the frame-to-frame pixel intensity differences of a representative excerpt of frames (4000 frames). Subsequently, frame-to-frame motion of all frames was projected into the first 500 principal components. Total video motion energy (video ME) was taken as the absolute sum across all 500 components. To investigate the relationship between neural measurements and more detailed orofacial movements, the first *n* principal

components were selected that captured movements explaining most of the frame-to-frame pixel intensity differences. For the regression model this was n=25 PCs, and for population feature decoding this was n=30 PCs. The first 25 PCs captured roughly 62% of the variance in frame-to-frame pixel intensity differences.
Pupil size and position were extracted using DeepLabCut (Mathis et al., 2018). The network was trained on 300 frames from 15 video excerpts of 1-2 minutes with varying illumination, contrast, pupil size, imaging angle, and task conditions. We labeled the pupil center and 6 radially symmetric points on the edge of the pupil. An ellipsoid was fit to these 6 outer points. The x and y coordinates of pupil center were taken as the center of the ellipsoid and pupil area as the ellipsoid area from the fitted ellipse parameters. Poorly fitted frames (likelihood <0.9999, output DeepLabCut) were replaced by the running median (median of 10 good frames), except if more than five adjacent frames were poorly fit (e.g. during extended periods of eye closure). We z-scored the total session traces.

Behavioral dominance
Behavioral dominance was quantified per trial condition by computing a behavioral dominance index (BDI):

$$BDI = \frac{A_{lick} - V_{lick}}{A_{lick} + V_{lick}} \qquad (Eq. 4.2)$$

where A_{lick} and V_{lick} are the amount of conflict trials in which the animal chose the auditory or visual lick spout, respectively. Note that misses are not taken into account in this index. BDI values range from +1 to -1, where +1 means exclusively auditory choices and -1 exclusively visual ones. The saliency-matched BDI (smBDI) was obtained by averaging the BDI of saliency-matched conflict trials ($A_{sub} + V_{sub}$, $A_{thr} + V_{thr}$, etc.). To describe and determine behavioral dominance as a function of audiovisual SOA we fitted the behavioral data with a cumulative Gaussian function:

$$BDI(t) = 1 - \gamma + (2 - \gamma - \lambda)\left(-\frac{1}{2}\left[1 + erf\left(\frac{t - \mu}{\sigma\sqrt{2}}\right)\right]\right) \qquad (Eq. 4.3)$$

Here, γ is the asymptotic visual dominance, λ the asymptotic auditory dominance, μ the mean of the cumulative Gaussian (time point of crossover), and σ the standard deviation (sensitivity to variations in SOA). We fitted the data using MATLAB's *fit* function and constrained μ between -300 and +300 ms, σ between 0 and 200 ms, and γ and λ between 0 and 1. Bootstrapped 95% CI was computed from n=1000 fits on resampled data.

Neural data processing
Before spike sorting the median of the raw trace of nearby channels (within 400 μm) was subtracted to remove common noise artefacts. Automatic and manual spike sorting were done using Klusta and the Phy GUI, respectively (Rossant et al., 2016). During manual curation each putative single unit was inspected based on its waveform, autocorrelation function, and its firing pattern across channels and

time. High-quality single units were included as having (1) an isolation distance higher than 10 (Schmitzer-Torbert et al., 2005) (2) less than 0.1% of their spikes within the refractory period of 1.5 ms (Bos et al., 2017; Vinck et al., 2016), (3) stable presence throughout the session. This latter was quantified by binning the firing across the entire session (approximately 45-75 minutes) in 100 time bins and only including neurons that spiked in more than 90 time bins.

To compute firing rates, spikes were binned in 10 ms bins and convolved with a causal half-Gaussian window with 50 ms standard deviation, unless stated otherwise. For analyses where neurons were compared, the firing rate of each single unit was z-scored by subtracting for each trial the mean firing rate of the baseline period (-1 to -0.1 seconds before stimulus change) and dividing by the standard deviation of all baseline periods.

For the initial assessment of how many neurons were significantly modulated after visual or auditory stimulus changes, the firing rate during baseline (-1000 to 0 ms) and post-change window (0-200 ms post stimulus) was compared with a paired two-tailed Wilcoxon signed rank test ($p<0.025$). Neurons were deemed significantly visually responsive if the firing rate was significantly different for at least one of the two grouped orientations (i.e. A/B or C/D) during maximal saliency trials (and similarly for auditory trials). Only conditions with at least 10 trials were tested. The fraction of significantly responsive neurons was only computed for sessions with at least 15 simultaneously recorded V1 neurons.

The onset latency of spiking activity for individual neurons was estimated using ZETA, a recently developed bin-less statistical test for determining whether a neuron shows a time-locked modulation of spiking activity (Montijn et al., 2021). We opted for this as visual and auditory stimuli can elicit very different neural dynamics in visual and auditory cortex (specifically, spiking responses in A1 can be very brief (DeWeese et al., 2003)) and ZETA prevents confounds related to different temporal dynamics by avoiding the need to bin spikes. ZETA was computed over time for auditory and visual maximal saliency trials. Neurons were deemed significantly modulated if ZETA exceeded a value of 2 during 0-1000 ms after stimulus change. The onset of this spiking response was taken as the onset latency. For Ext. Data Fig. 4.7g, we focused on sensory-evoked spiking and to minimize occlusion by motor-related confounds, we excluded auditory and visual spiking activity with onsets occurring later than 200 ms (note, however, that this approach does not strictly separate sensory from motor-related activity).

To estimate the onset latency of visually or auditory induced spiking activity with greater temporal detail, the spiking activity was pooled across neurons in an area. Only sessions with at least 10 neurons were included. The spike train of each neuron was divided across 1 ms bins and smoothed with a causal half-Gaussian window with 10 ms standard deviation. The activity was averaged over trials of interest (Vmax or Amax trials). To compare across sessions, the firing rate was averaged across all simultaneously recorded neurons in each area and z-scored, as described for single neurons. The first time bin this z-scored activity crossed a threshold of 1 was taken as the onset latency of the population activity.

Feature tuning

Feature tuning in the change detection task was assessed using ROC analysis (Green and Swets, 1966), which quantifies how well an external observer could discriminate between two sets of values. The area under the ROC curve (AUC) was computed for the distributions of either the firing rate response of V1 neurons or video ME (0-200 ms) between grating orientations A/B or C/D, or auditory frequencies A/B or C/D. Each class had to have at least 10 trials. AUC values are in the range of 0 to 1, but were rescaled between -1 and 1, where -1 indicates complete selectivity to A/B and 1 to C/D.

Orientation and frequency tuning in the stimulus detection paradigm (Ext. Data Fig. 4.4) was assessed using the global Orientation Selectivity Index (gOSI). This measure adequately captures tuning in a circular domain (Ringach et al., 2002b). As the stimulus set in both the visual and auditory domain was circular (visual orientations, and auditory tones due to the Shepard harmonic weights, see above) this measure captured selectivity to stimuli in both modalities similarly. The gOSI was computed as:

$$g = \frac{\left\| \sum_{\theta} R(\theta) e^{2i\theta} \right\|}{\sum_{\theta} R(\theta)} \qquad (E\ .4.4)$$

Here $R(\theta)$ is the firing rate response of a neuron (0-200 ms) to either a grating moving along direction θ or a Shepard tone with center frequency θ and i is the imaginary unit. gOSI varies between 0 and 1, with 0 indicating a neuron completely untuned, and 1 a neuron only responding to a single orientation/frequency. Neurons were deemed significantly tuned if their gOSI exceeded 95% of the shuffled distribution (recomputing gOSI for n=1000 shuffles of orientation or frequency labels). Signal correlations were computed as the Pearson correlation of the trial-averaged tuning curve between pairwise tuned neurons.

Local field potential (LFP) analyses

The LFP was obtained by down-sampling the recorded voltage signal over time from 32000 Hz to 1024 Hz and low-pass filtered below 300 Hz (4^{th} order Butterworth filter). For current source density (CSD) and event-related potential (ERP) analyses the signal was further low-pass filtered below 100 Hz (4^{th} order Butterworth filter). The CSD profile was computed by applying standard Nicholson-Freeman calculations on the LFP signal with Vaknin transform (Vaknin et al., 1988) with 0.4 Siemens per meter as conductivity (Logothetis et al., 2007). We calculated the CSD profile for each of the linear arrays of electrodes on our polytrode configuration separately, interpolated between sites, and then merged the profiles. The ERP was the stimulus-onset locked trial-average LFP response.

To separate the sensory and motor contributions, trials were split into 'still' and 'moving' trials based on the amount of motor activity. The z-scored video ME was computed (0-500 ms post-stimulus change) and 'still' trials had z-scored video ME between -0.5 and 0.5 and 'moving' trials a z-scored video ME larger than 1. For

each trial type (e.g. Amax trials changing to set A/B) the CSD and ERP was computed for still and moving trials separately. Only conditions with at least 3 trials in both still and moving conditions were included. Subsequently, the different auditory and visual trial types (saliencies and features) were averaged within modality.
We excluded a subset of sessions in which movement artefacts were present (N=16 sessions excluded from 62 sessions with video and LFP recordings in V1). To identify movement artefacts, the ERP and the CSD were computed aligned to one lick event. Sessions with movement artefacts were easily identified by the presence of low-frequency large deflections in the LFP that were several fold larger than spontaneous activity, strikingly dissimilar between adjacent channels instead of smoothly varying across cortical depth. These sessions were not included in LFP analyses.

Cortical depth estimation
The laminar depth of each silicon probe in V1 was estimated based on a combination of two factors. First, we computed the CSD profile to contrast-reversing checkerboard stimuli. Before each session, we displayed full-field contrast-reversing checkerboards (full contrast, spatial frequency = 10 retinal degrees, temporal frequency of contrast reversal = 0.5 Hz, n=10 reversals). The earliest visible current sink was taken to indicate layer 4 (Niell and Stryker, 2008; Schnabel et al., 2018). Second, we computed the power in the 500-5000 Hz range of the raw, unfiltered signal for each channel (Senzai et al., 2019) and set the channel with highest MUA spiking power as the center of L5 at 600 μm from the dura and rereferenced all channels to this depth. Channels that were above 0 μm or below 1000 μm were excluded from the analyses.

Regression model
To quantify single neuron encoding of sensory and motor variables we constructed a linear regression model. This approach is particularly useful to disentangle the time-dependent contribution of experimenter-controlled task events and self-timed behavioral events to single-trial neuron firing rate. The model was trained to predict the firing rate (-500 to +1000 ms relative to stimulus change in 20 ms time bins, convolved with a gaussian with a standard deviation of 25 ms). We included four sets of predictors: trial number, visual, auditory and motor variables. The trial number predictor consisted of a whole-trial value scaled by trial number within that session. For sensory variables a separate predictor set was made per combination of orientation (or frequency) and amount of change, taking simultaneously into account the selectivity of neurons for features and saliency. For a given stimulus, there was a separate predictor for each post-stimulus time bin (50 time bins from 0 to 1000 ms). This resulted in 50 time bins x 2 saliencies (thr and max) x 2 stimuli (set A/B and set C/D) = 200 predictors per modality. For motor variables the first 25 video PCs were included. For convenience, all

predictors were normalized to their maximum values before being fed into the model.
This resulted in a predictor matrix of size P x T for each neuron, where P is the number of predictors (1 trial number, 200 visual, 200 auditory, and 25 video predictors = 426 predictors) and T is the number of total time bins. The regression model was fit on concatenated single trials and T is therefore the number of trials (typically 200-500 trials per session) multiplied by the number of time bins per trial (75 time bins; -0.5 to +1 sec relative to stimulus change, 20 ms time bins).
The model was fit on catch, auditory, and visual trials (audiovisual trials were excluded during fitting) from all sessions with V1 or AC recordings during the change detection task with recorded video and without pharmacological or lick-spout manipulations. Sparsely firing neurons produced fitting difficulties and neurons with a session-average firing rate below 0.5 Hz were excluded. The total dataset for the regression analysis consisted of: n=51 sessions, NE: 9, UST: 10, MST: 32 sessions, 19217 trials, 790 V1 and 99 AC neurons. For analyses of conflict trials, the model was fit on trials excluding conflict trials and tested on the held-out conflict trials. As conflict trials were only presented in a subset of sessions and not in UST mice this dataset consisted of 37 sessions (NE: 9; MST: 28 sessions, 16021 trials, 648 V1 neurons). The model was fit with a Gaussian link function with the *glmnet* package in Matlab (Friedman et al., 2010). We used elastic-net regularization ($\alpha = 0.95$) and 5-fold cross-validation. To maximally punish weights without losing model fit quality, lambda was maximized while minimizing the cross-validated error. We quantified model performance by assessing the 5-fold cross-validated explained variance (EV):

$$E = 1 - \frac{\mathrm{var}(Y - \hat{Y})}{\mathrm{var}(Y)} \qquad (E.4.5)$$

where Y is the original firing rate and $\hat{Y}$ the predicted firing rate. EV was computed for the concatenated firing rate during a specified time window (0-200 ms for Fig. 4.3 and 4.4; and 0-500 ms for Fig.4.5).
After fitting, the model could be used to predict firing rates on held-out test trials or with restricted predictors. To estimate the contribution of different predictors to firing rate variability, EV was computed using only one set of predictors (all other predictors were set to zero). Regularization and cross-validation already minimized overfitting to predictors, but to further verify that predictors were not capturing unrelated variance, we fit the model with one set of predictors circularly shuffled across time within the session. Thus, the original temporal relationship between for example auditory stimulus predictors and actual auditory-evoked activity was destroyed. The additional EV explained by the intact model relative to the shuffled model was taken as uniquely explained variance (Musall et al., 2019). We found very similar results for a linear model that used smooth temporal basis functions instead of boxcar bins, or when using different elastic net mixing parameters. Note that our measure of predicting single trial binned spike counts

leads to low levels of explained variance (Steinmetz et al., 2019), while we obtained high levels of explained variance when predicting trial-averaged activity (Runyan et al., 2017). Here, we were however interested in explaining single trial firing rate due to trial-by-trial differences in orofacial movements.

Multivariate stimulus decoding from neural and video data

We used multivariate analyses to decode visual orientation or auditory frequency from either V1 population spiking activity or dominant orofacial movements (video PCs). In all decoding analyses, nearby orientations/frequencies were grouped together (denoted as set A/B versus C/D) in order to have a two-class classification problem, and here we considered only large stimulus changes (Amax and Vmax). Only sessions with V1 recordings and at least 15 trials for each orientation/frequency were included. We balanced the two classes with random subsampling of the majority class.

For the analysis in Figure 2j-o we decoded auditory or visual feature identity from simultaneously acquired data from individual sessions. We subselected all sessions which contained at least 5 neurons recorded in V1. Spikes were binned in 200 ms time bins and advanced by 25 ms. For orientation/frequency decoding using video footage, 'population' data was created by replacing the binned spike counts with the binned values of the first 30 principle components of the motion energy (video PCs).

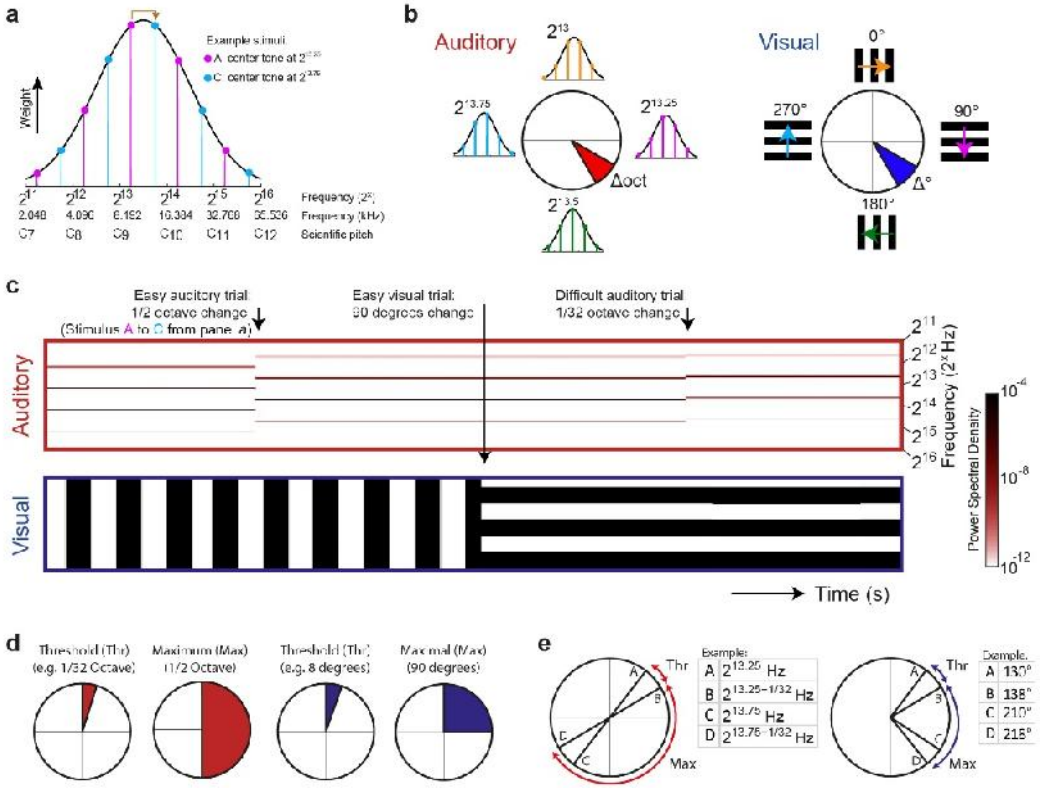
We trained a support vector machine (SVM; linear kernel) using stochastic gradient descent (as implemented in scikit-learn (Pedregosa et al., 2011)) to predict the orientation/frequency at every time point. We employed 3 repeats of a 3-fold stratified cross-validation routine, whereby trials for training and testing are drawn randomly, but the equal ratio of the two classes is preserved in each set. Note that the same train/test splits are used across all time points in a given bootstrap iteration. Features (neuronal spike counts or video PCs) were standardized to have zero mean and unit variance (features of the test set were standardized using the mean and standard deviation of the training set). Reported decoding performance is the accuracy on the held-out test data. The average decoding accuracy averaged across the time bins whose edges did not exceed the 0-300 ms range were used to generate the scatter plots.

For the analyses in Figure 5g-j pseudopopulation data was constructed by combining data acquired during different sessions. We employed a bootstrapping procedure in which at every iteration we randomly sampled a subset of V1 neurons or video PCs across all recording sessions. When training the model on unisensory trials and testing on audiovisual trials (or vice versa), we selected only sessions in which for each trial type there were at least 10 trials for each orientation/frequency. When training and testing the model on the same trial type, we required at least 20 trials for each orientation/frequency. For every bootstrap iteration, we constructed a train and a test set by randomly sampling for every session 10 trials of each class for the train set, and 10 trials of each class for the test set (in the case where we trained and tested on the same trial type, trials appearing

in the training set did not appear in the test set). Spikes were binned in 100 ms time bins and advanced by 25 ms. For every time point, we then assembled a feature matrix *X_t* of size 20 (2x10) by N, where N is the number of neurons, such that column *i* is the binned spike count of neuron *i* at time t in the 20 subsampled trials, and row *j* is the binned spike count at time *t* of all randomly selected neurons of a pseudo-trial that merges data from different sessions. For orientation/frequency decoding using video footage, pseudopopulation data was created in the same way by replacing the binned spike counts with the binned values of the first 30 principle components of the motion energy (video PCs). In other words, a similar data matrix was constructed subsampling video PCs (from the first 30 video PCs) during trials of specific orientation/frequency from different sessions.
The SVM was then fitted on the train set and evaluated on the test set at every time point (without a full cross-validation routine). The SVM coefficients (Fig. 4.5k) were obtained with an adaptation of the pseudopopulation method. For every bootstrap iteration, a random subset of 50 V1 neurons was selected. Using 3x3 stratified cross-validation routine, a linear SVM was fitted on the train set (consisting of either visual or audiovisual trials), and the coefficient of each neuron was recorded. The same pseudopopulation approach was also employed in Fig. 4.5g, where the orientation/frequency label was replaced by the trial type (visual versus auditory).

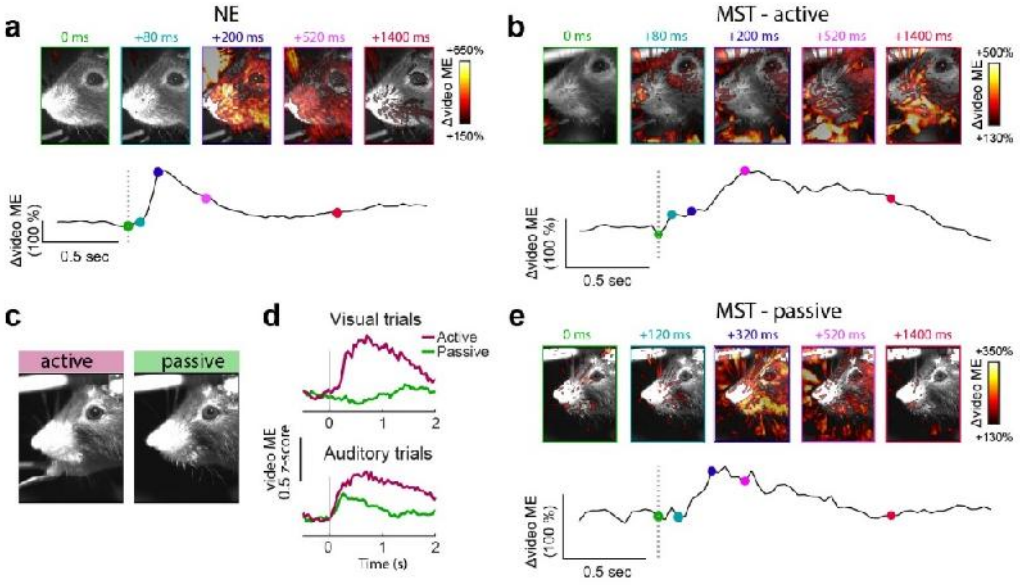
Statistical analysis
Unless specified otherwise, all statistics were performed using linear mixed models (LMMs) in MatLab (MathWorks, Natick, MA). LMMs can account for the hierarchical nature of our data (neurons and trials sampled from the same mice)(Aarts et al., 2014). LMMs describe the relationship between a response variable and multiple explanatory variables, and comprise two types of explanatory terms. Fixed effects are the variables of interest, while random effects, also commonly referred to as “grouping variables”, specify and account for the group. For all analysis involving hierarchical data, LMMs were constructed with mouse identity as a random effect (intercept only). Importantly, mouse identity was not included as a random effect for analyses with cohort as fixed effect, as variability between mice was key to those results. Statistical tests were performed on the fixed effect using ANOVAs on the LMMs. To estimate the denominator degrees of freedom (DF2) for F-tests, the Satterthwaite approximation was used for LMMs. Linear hypothesis tests were performed in the case of posthoc comparisons using the relevant contrasts. Non-nested data was tested using nonparametric methods. Results with a p-value lower than 0.05 were considered significant. When multiple, independent comparisons were performed, p-values were corrected by applying a Bonferroni correction.

Extended Data

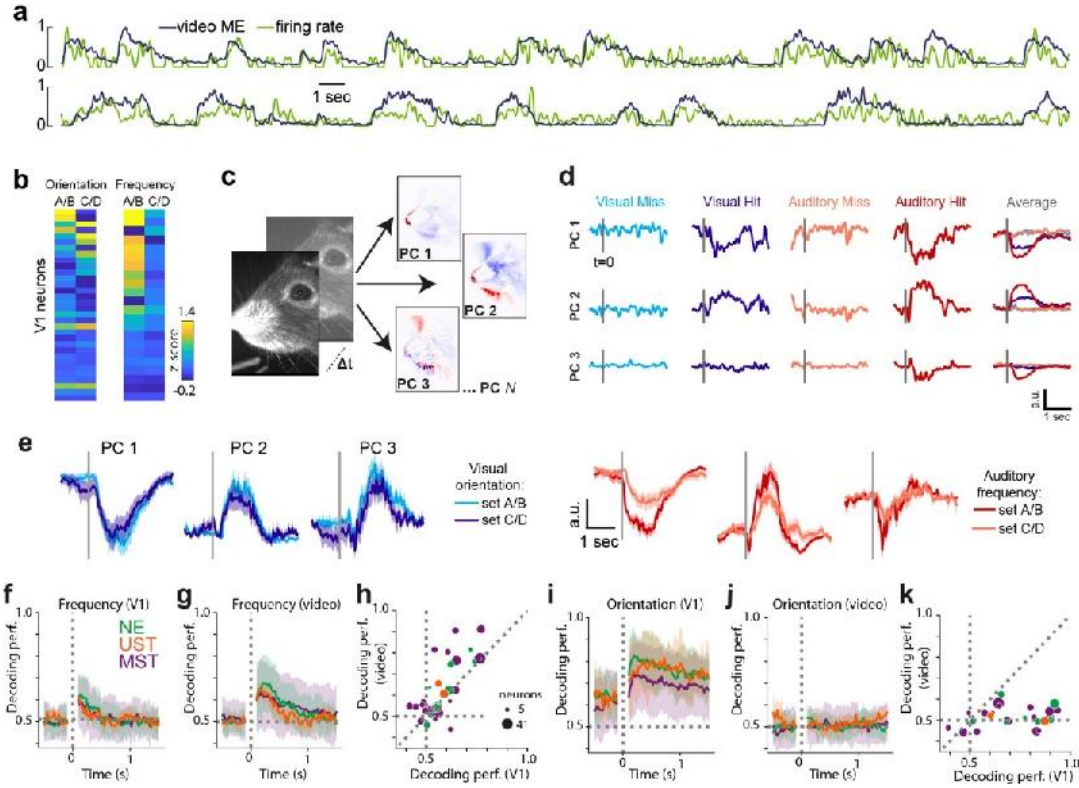


Extended Data Figure 4.1: Details of auditory and visual stimulus design. a) Each auditory stimulus was composed of five pure tones at harmonic frequencies (octaves below and above other tones). The weight with which each tone contributed to the overall stimulus was taken from a Gaussian distribution across all possible tones. The example stimulus A in pink is composed of a tone of $2^{13.25}$ Hz (center tone, highest weight) and two lower (at $2^{11.25}$ and $2^{12.25}$ Hz) and two higher harmonics (at $2^{14.25}$ and $2^{15.25}$ Hz). Tones followed scientific pitch and are expressed as powers of two: 2^{13} corresponds to 8.192 kHz, and C_9 in scientific pitch notation. During an auditory trial, the stimulus changed to a stimulus of five new harmonic tones with different weights (for example stimulus A to B). **b)** The left polar diagram shows the circular arrangement of auditory stimuli. For each cardinal direction the insets show the tonal weights associated with these stimuli. Note how ever increasing the center tone frequency ultimately results in a circular shift back to the starting stimulus. This circularity can also be seen in panel *a*: going up and down half an octave from stimulus A always results in stimulus B. The auditory stimulus set is therefore circular. This feature is exploited in the Shepard illusion of eternal rise or drop in pitch. However, our stimuli were static so the illusory effect of continuously increasing or decreasing pitch was absent. The only illusory component was that half an octave change could be both experienced as an increase or decrease in pitch. This circular design of auditory stimuli mirrors the visual stimulus set (right part) with drifting gratings in all orientations. The amount of frequency change (expressed in partial octaves, red) or orientation change (expressed in degrees, blue) determined the saliency of auditory and visual changes. **c)** Example stimuli during three consecutive trials. The upper spectrogram over time includes two auditory change trials. Auditory stimuli continued to be presented until the next auditory change, which could be identified

based on a difference in spectral content, and experienced as a change in pitch. The example shows an easy auditory trial (salient change; stimulus A to B, half an octave) followed later by a difficult trial (subtle change; 1/32 of an octave). The lower schematic shows visual orientation over time including a visual trial. Note that the gratings were continuously drifting in the direction orthogonal to the grating orientation. An audiovisual trial would consist of a simultaneous change in both modalities (not shown). Note that this is only a schematic depiction, hence time is depicted in arbitrary units. **d)** Schematic of the different levels of saliency (i.e. amount of change) between threshold and maximal saliency. Threshold saliency was titrated per mouse based on task performance. **e)** The stimulus set during recording sessions was limited to four visual and four auditory stimuli with two levels of change between them. Tables show one example stimulus set for each modality, but stimuli were varied across sessions.

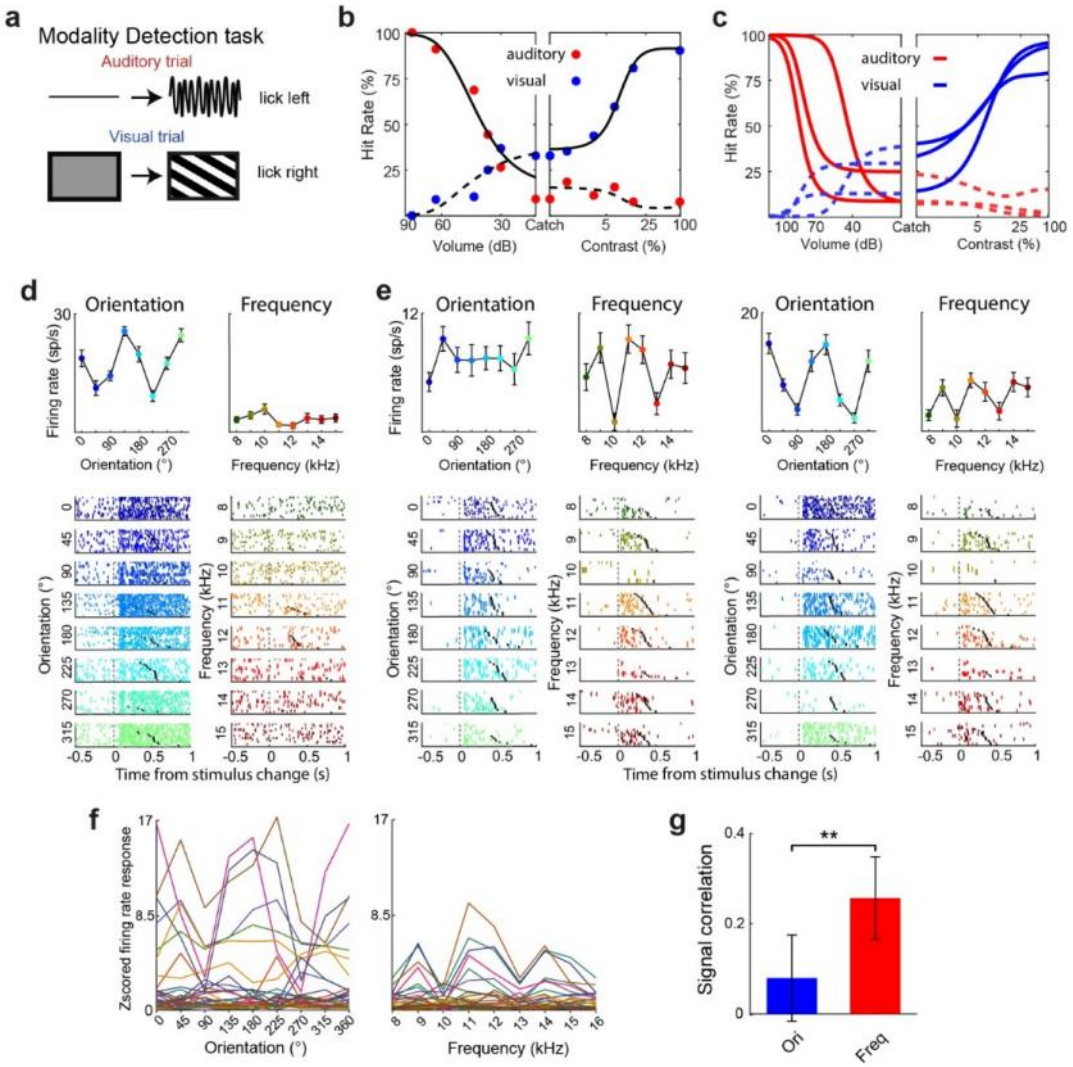


Extended Data Figure 4.2: Sounds evoke instructed and uninstructed orofacial movements. a) Sounds evoke brief whisking and eye twitching movements in NE mice (example session). Upper images show heatmap of the increase in video ME overlaid on one reference frame. Lower trace shows video ME averaged over auditory trials with dots highlighting time points of upper frames. **b)** Same as b, but for an example MST session. Here auditory trials not only evoked whisking and eye movements (uninstructed), but also continued instrumental licking movements as mice were rewarded for reporting auditory stimuli. **c)** To further test whether the increase in motor activity was not associated with licking behavior, we continued sensory stimuli but removed the lick spout. Blocks of active trials (with lick spout, left image) and passive trials (without lick spout) were interleaved during a session. **d)** The increase in video ME normally seen following visual stimuli (due to report-related licking movements) was absent during passive blocks. On the other hand, auditory stimuli continued to evoke orofacial movements during passive blocks in the absence of licking to a rewarded lick spout. These results are in line with the comparison between cohorts (Fig. 4.2b) where unrewarded auditory stimuli (but not unrewarded visual stimuli) still evoke orofacial movements. Motor-related confounds are thus important to control for not only in auditory behavioral tasks, but also naive animals. **e)** Same as (a, b), but for auditory trials during passive blocks of an example MST session. Auditory trials continued to evoke uninstructed orofacial movements, but less prolonged due to the absence of licking movements.



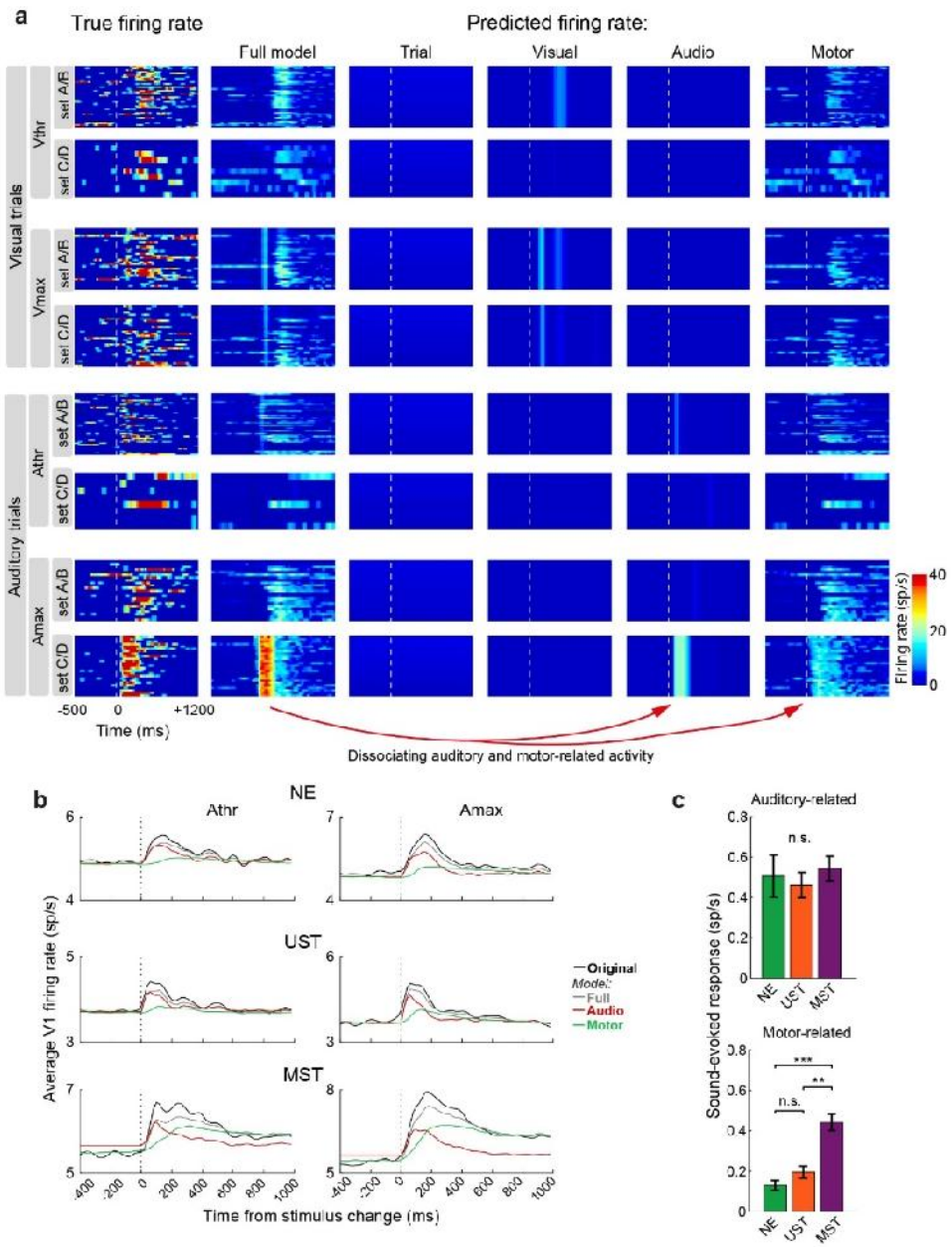
Extended Data Figure 4.3: Detailed orofacial movements underlie frequency-tuned activity. a) Normalized firing rate and video ME over time for two example V1 neurons. Top: r=0.61, UST mouse. Bottom: r=0.55, MST mouse.
b) Activity heatmap for simultaneously recorded V1 neurons showing a distribution of selectivity to orientations but similar tuning to auditory frequency. Left and right are taken from different sessions. **c)** To extract more detailed video information, we applied PCA to the frame-to-frame pixel intensity difference (FaceMap; Stringer et al., 2019) and extracted principal components that captured the most dominant movements (video PCs). Most movement was confined to snout, whisker pad and tongue regions. PC: principal component. **d)** Example traces of the first three PCs during individual trials of different modalities and decisions. Hit trials were associated with motor activity during lick responses and reward consumption. Data from one MST session. Gray line indicates stimulus change. **e)** First three video PCs for an example session showing similar movements following changes in visual grating orientation, but variable movements following different auditory stimuli. V1 could still encode auditory features beyond what is explained by the modulatory effects of orofacial movements. We therefore tested how well we could decode stimulus identity by considering population spiking activity in V1, and compared this to detailed video analysis. A population decoder (support vector machine, SVM) was trained to discriminate auditory or visual stimulus identity using either the spiking data or these video PC values. **f)** Auditory stimulus frequency could be decoded from V1 population activity. Decoding performance of decoders trained to discriminate post-change auditory frequency from V1 population activity. Horizontal dashed line indicates chance level. Line and shading indicate mean and 95% CI. **g)** Same as (f), but for decoding auditory frequency from the first 30 video PCs. Auditory stimulus frequency could be decoded from video data. **h)** Relationship across sessions between auditory frequency decoding performance using V1 data (x-axis) and video data (y-axis). Decoding performance was highly variable across sessions and, interestingly, strongly

correlated between spiking and motor activity ($R=0.71$, $F(1,17)=29.13$, $p=4.49*10^{-5}$). Those sessions with frequency-selective orofacial movements thus also displayed frequency-selective population activity. Further, video decoding outperformed neural decoding ($F(1,49)=8.25$, $p=0.006$). Dot size scales with number of simultaneously recorded neurons for that session and dot color indicates cohort. **i)** Same as (f), but for decoders to discriminate post-change visual orientation. Visual grating orientation could be decoded from V1 population activity. Baseline coding results from the fact that gratings jumped between the same stimuli (A/B to C/D and vice versa) and neurons showed persistent selectivity, seen in (Fig. 4.2h). **j)** Same as (i), but for visual orientation. Visual grating orientation could not be decoded from video PCs. **k)** Same as (h), but for the relationship between orientation decoding based on V1 activity versus video PCs. Decoding performance was not correlated across sessions ($R=0.15$,$F(1,28)=0.67$, $p=0.42$) and higher for V1 spikes than for video PCs ($F(1,42)=51.14$, $p=9.09*10^{-9}$). Although absolute decoding performance from these qualitatively different sources is less meaningful, the dissimilarity between modalities is striking.



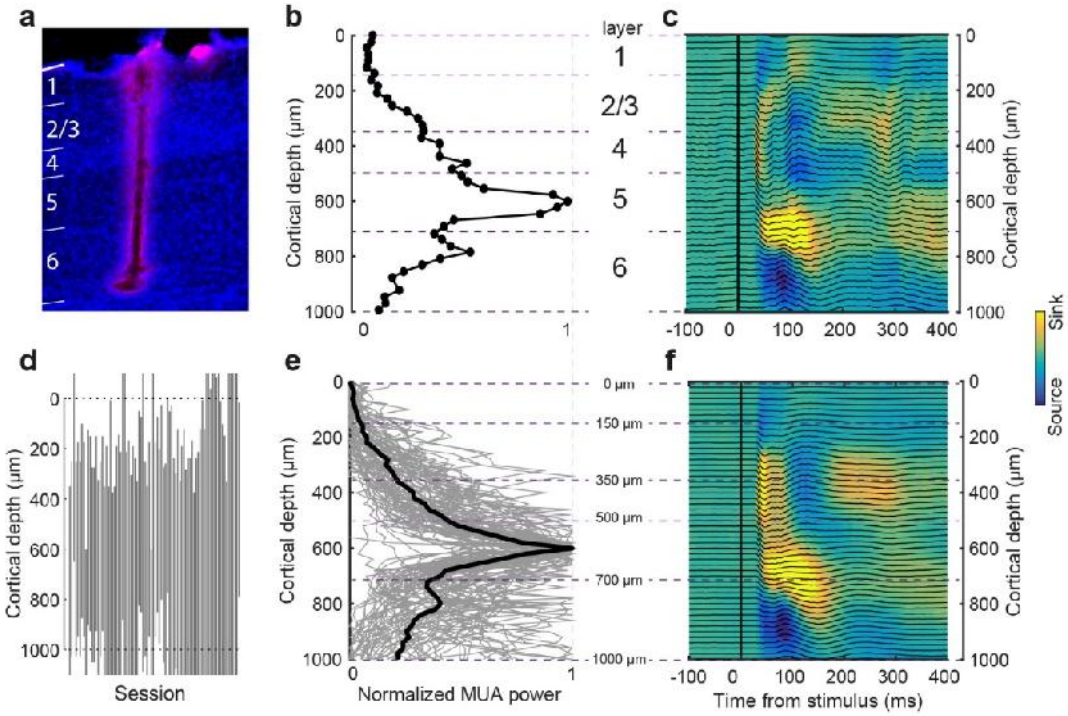
Extended Data Figure 4.4: Similar frequency tuning of primary visual cortical neurons during audiovisual stimulus detection. a) To establish whether our findings generalized beyond our change detection task, we trained animals (n=3) to detect the presence of auditory and visual stimuli (same stimulus set as in the change detection task) and to discriminate and selectively report the modality, as in the MST task of our main change detection paradigm. Rewards were allocated upon licking to the auditory lick spout after the onset of one of eight tones, and upon licking to the visual lick spout to one of eight gratings was rewarded. **b)** Performance on an example session on the detection of auditory stimuli of varying volume (left panel) and of varying contrast (right panel). Note how auditory and visual hit rates increase as a function of volume and contrast, respectively. The behavioral data was fit with the same two-alternative signal detection model as behavioral data from the change detection task. Behavioral response rates are shown as dots, model fits as lines. **c)** Average psychometric fits for each mouse obtained by averaging the parameters of single session fits. **d)** Raster plot and tuning curve of an example orientation-tuned V1 neuron. Upper panels show firing rate (0-200 ms) in response to eight drifting grating orientations (left) and eight compound Shepard tones

with center tone spaced between 8 and 15 kHz (right). Dot and error bar show mean ± SEM across trials. Colored tickmarks in the lower raster plots show trial-by-trial spiking. Black tick marks indicate first lick after the stimulus. Note the classical orientation tuning expected from V1 neurons in response to full-field oriented drifting gratings. Auditory frequency tuning was not significant. **e)** Same as (d), but for two V1 neurons from the same session where the auditory response depended on the frequency components of the auditory stimulus. Note how the neurons are similarly tuned and their firing rates are associated with licking behavior as well. **f)** Tuning curves for orientation and frequency for all V1 neurons (individual lines) from one session. Note dissimilarity in orientation tuning, but similarity in frequency tuning. **g)** The signal correlation of all significantly orientation-tuned (left) and frequency-tuned (right) neurons. Signal correlations were computed as the Pearson correlation of trial-averaged tuning curves between neuronal pairs. Signal correlation was higher between frequency-tuned neurons than orientation-tuned neurons ($F(1,406)=9.50$, $p=0.0022$; n=148 signal correlations from 23 orientation-tuned V1 neurons, n=258 from 36 frequency-tuned V1 neurons). The finding that V1 neurons responded to the same frequencies (those associated with motor movement, Fig. 4.2i; Ext. Data Fig. 4.3b) suggests that variability in motor variables drives tuning. $^{**}p<0.01$.

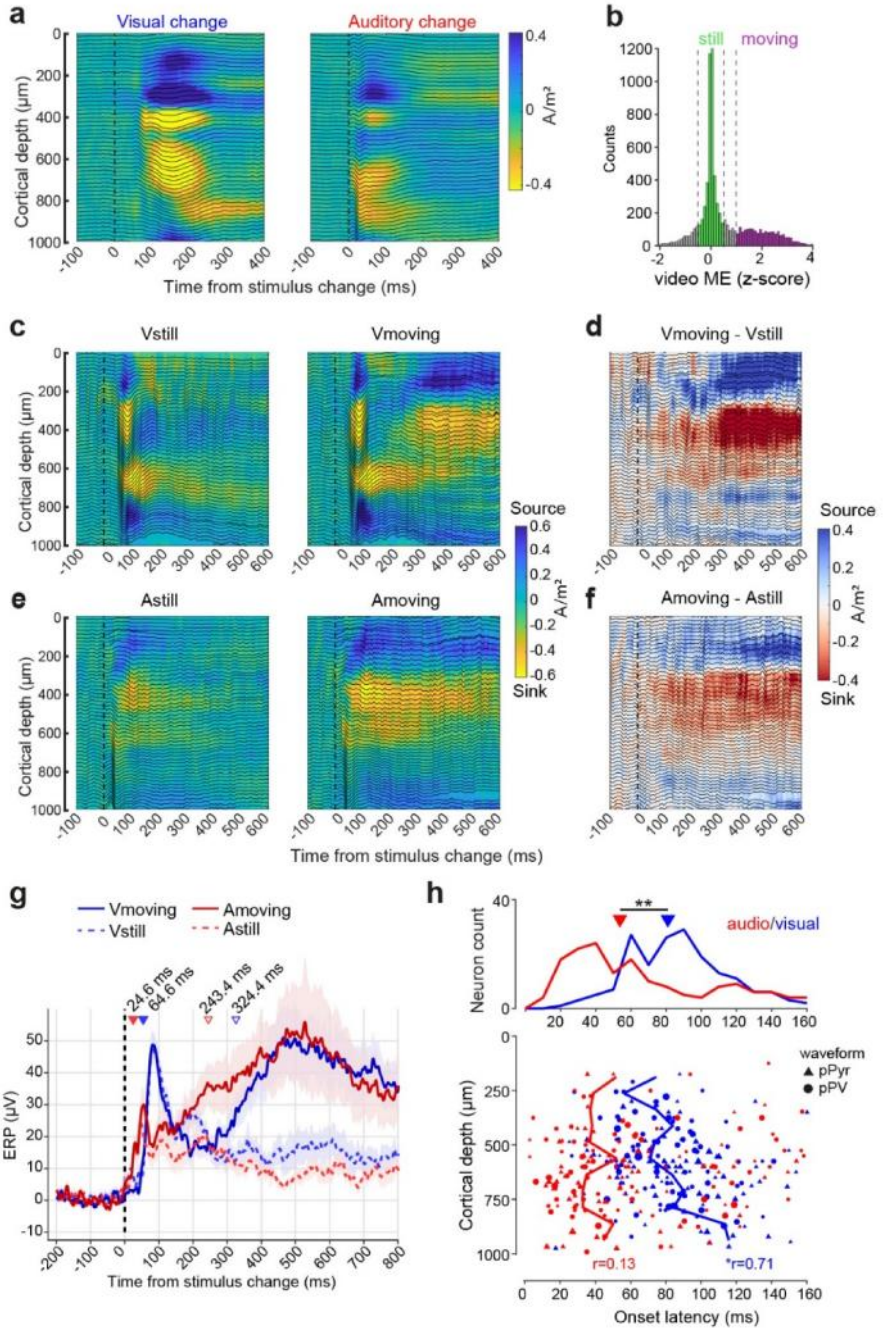


Extended Data Figure 4.5: Dissociating visual, auditory and motor-related activity using a regression model. a) Each heatmap shows the firing rate over time for a subset of trials with each row representing a different trial type, and each column a different source of the firing rate. The leftmost column shows the original firing rate. The second column shows the predicted firing rate for the same trials using all predictors in the model. The remaining columns show the predicted firing rate using only a subset of the predictors. For this example neuron, the trial number explained little variability (trial number captured response drift across the session for some other neurons, not shown). Visual

predictors explained an early response transient especially in Vmax trials. Auditory predictors captured an early response transient in some auditory trials (set C/D), whereas motor variables (the first 25 video PCs) captured variability across visual and auditory trial types. **b)** Same as Figure 4.3f, but for each of the task cohorts separately and auditory trials only. Auditory-related activity was present in all three cohorts. Sound-evoked motor-related activity was larger in the MST cohort, quantified in (c). **c)** Predicted sound-evoked response (0-200 ms minus baseline activity) for each of the cohorts using either auditory predictors (top) or motor predictors only (bottom). Cohorts did not significantly differ in auditory-related activity ($F(2,790)=0.18$, $p=0.835$), while motor-related activity was significantly different ($F(2,790)=8.07$, $p=0.00034$) and significantly larger in MST mice compared to NE and UST (Posthoc comparison: NE vs. MST: $F(1,787)=11.4$, $p=0.000789$; UST vs. MST: $F(1,787)=7.3$, $p=0.00702$; NE vs. UST: $F(1,787)=0.3$, $p=0.582$). The larger sound-evoked response in MST mice in Figure 4.1e,f is therefore attributable to increased motor-related and not auditory-related activity. Mean ± SEM across neurons.

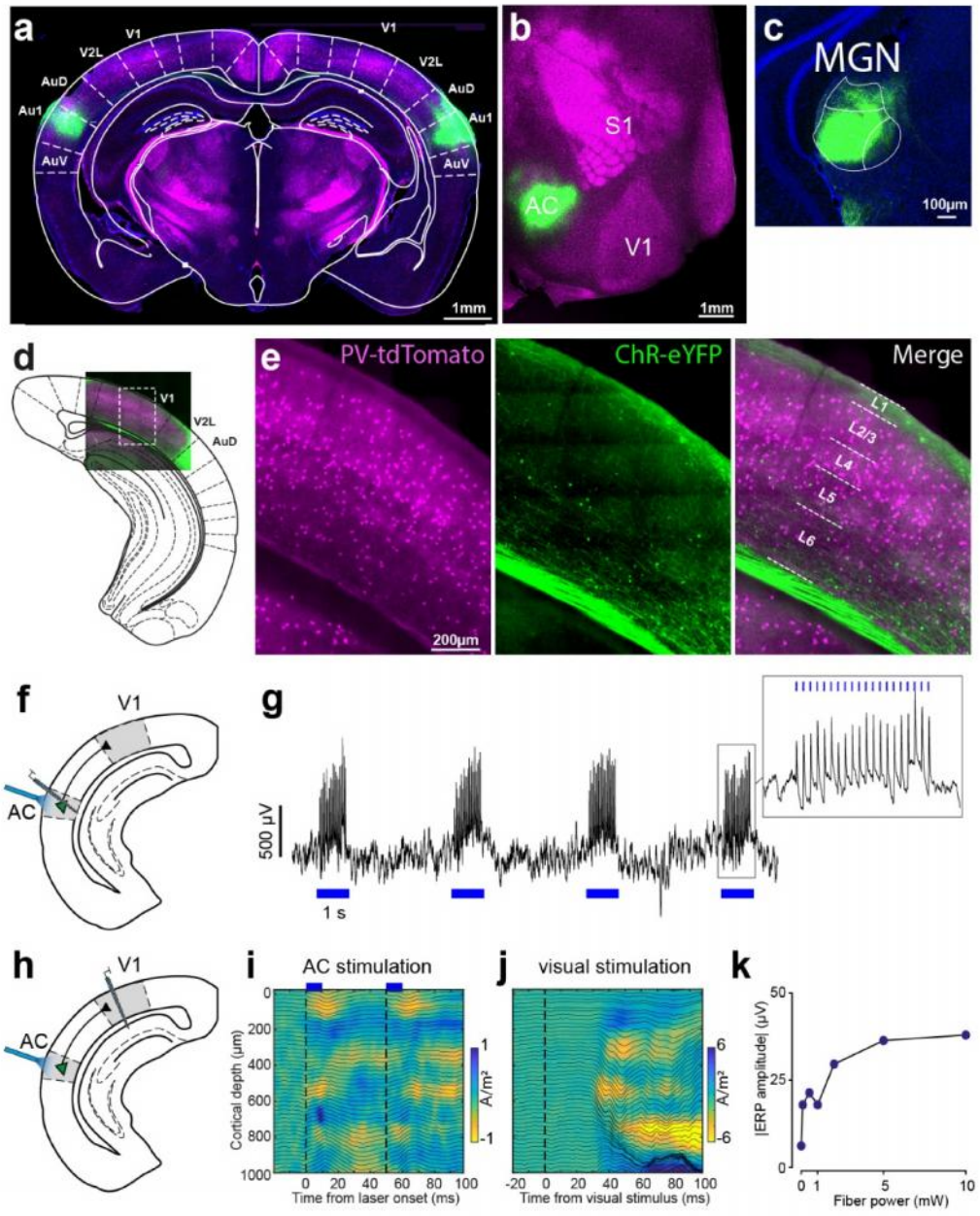


Extended Data Figure 4.6: Cortical depth estimation in primary visual cortex using physiological markers. a) Close-up of a coronal section on V1 showing the electrode track stained with DiI. **b)** Example distribution along the probe of spectral power (500 Hz to 5 kHz) indicative of multi-unit activity (MUA). High MUA power is characteristic of L5. Compare with (Senzai et al., 2019). **c)** Current source density (CSD) map and LFP traces (black lines) in response to checkerboard stimulation. *B* and *C* are from the same example session. Color corresponds to CSD power. **d)** Overview of electrode span across layers. Each line is one session (n=84 sessions). Data from electrodes at depths above 0 or below 1 000 µm were excluded from analyses. **e)** Same as b, but for all sessions. Gray lines are individual sessions, black line the median. **f)** Same as c, but averaged across all sessions.



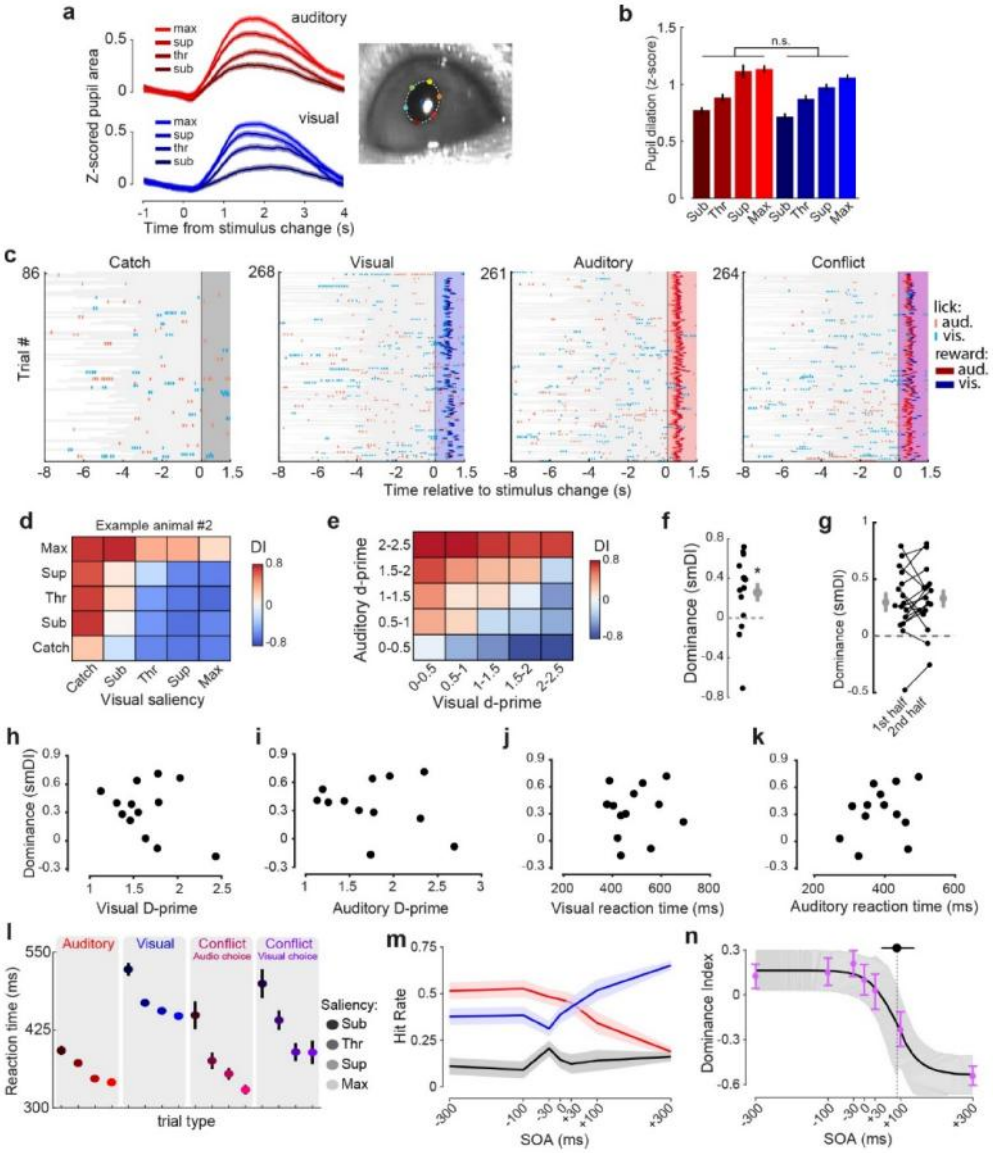
Extended Data Figure 4.7: Early sensory and late motor-related components of current-source density and cell-type spiking profiles in visual cortex. a) The current source density (colormap, CSD) and event-related potential (black traces, ERP) for auditory and visual stimulus changes in the same example session (MST mouse). **b)** Histogram of z-scored video ME (0-500 ms post-change) across visual and auditory trials of all

sessions with LFP recordings in V1 (all cohorts). To separate the contribution of motor activity to the LFP, all trials were split into 'still' and 'moving' trials based on the amount of motor activity. 'Still' trials had z-scored video ME between -0.5 and 0.5 and 'moving' trials a z-scored video ME larger than 1. **c)** For each session a CSD map was constructed using either still or moving trials given the same visual stimuli. Average across n=46 sessions (NE: 12 sessions; UST: 7; MST: 27). Visual stimuli evoked a consistent and characteristic current source density (CSD) profile with an early sink in L4 and subsequent sink-source pairs in L2/3 and L5/6, in line with earlier reports (Niell and Stryker, 2008; Schnabel et al., 2018; Senzai et al., 2019). **d)** The difference between the Vstill and Vmoving maps in (c), which we interpret as mostly related to motor differences. Note how most of the motor-related CSD power is expressed after 200 ms in L2-5 and predominantly in superficial and middle layers. **e)** Same as (c), but for auditory trials. Note how the early sinks and sources in deep layers of the auditory CSD map in the example session of (a) are only partially reflected in the average. **f)** Difference map of the Astill and Amoving maps in (e). Note how the movement-associated CSD pattern resembles that of visual trials (d), but is generated somewhat earlier in time. **g)** Absolute ERP response (in μV) averaged across cortical depth for selected trial categories. The tick marks and text denote the first time bin the LFP response is different from baseline (-500 to 0 ms) during auditory or visual trials irrespective of motor activity ($p<0.05$, Wilcoxon signed rank test, Bonferroni correction). These latencies closely match spiking onset latencies (Fig. 4.3b). The LFP response for auditory trials can be seen to diverge between still and moving trials around 100 ms after stimulus onset and was significantly different after 243.4 ms (bootstrap test, n=1000 resamples, $p<0.05$) and after 324.4 ms for visual trials ($p<0.05$) suggestive of late motor-related signals. Line and shading are mean ± SEM. **h)** Laminar organization of onset latencies of visual and auditory responses in V1 (spiking data, not LFP). Top histogram shows the distribution of onset latencies of all significantly auditory responsive neurons (red) and visually responsive neurons (blue). Significance and onset latency were assessed using a binning-free algorithm, ZETA (Montijn et al., 2021). Spiking onset was significantly earlier for auditory versus visual stimuli (55.3 ms (31.4 - 108.5 ms) versus 80.3 ms (61.5 - 98.5 ms); median and interquartile range; $F(1,411)=5.37$, $p=0.0209$), similar to our earlier population-averaged approach (Fig. 4.3b). Bottom panel shows each neuron's onset latency as a function of its recorded depth and cell type. If neurons are bimodally responsive they appear twice. Symbols are scaled by response magnitude. Putative pyramidal cells (broad-spiking) and putative parvalbumin expressing cells (narrow-spiking) were classified based on their waveform. L1 is mostly empty because almost no cells were recorded in that layer. *$p<0.05$, **$p<0.01$. Visually driven cells first began to fire significantly in the middle and superficial layers and later in deeper layers, consistent with the canonical sensory processing scheme (Douglas and Martin, 2004; Harris and Shepherd, 2015). Auditory-evoked firing started at similar latencies across layers, with many auditory responsive neurons in deep layers. Cortical depth was significantly correlated to spiking onset latency during visual trials ($r=0.71$, $p=0.015$, Pearson correlation), but not auditory trials ($r=0.13$, $p=0.696$).



Extended Data Figure 4.8: Auditory cortical projections modulate superficial and deep layers of primary visual cortex. a) Coronal section showing bilateral AC expression of AAV2-CaMKIIa-hChR2(H134R)-eYFP (*green*: eYFP) in a PvCre-tdTomato mouse (magenta: tdTomato), centered at primary auditory cortex (Au1). AuV: ventral secondary auditory cortex. AuD: dorsal secondary auditory cortex. V2L: lateral secondary visual cortex. **b)** Same as a, but for a flattened cortical section showing ChR2-eYFP expression in AC. **c)** Close up of densely labeled projections in medial geniculate nucleus of the thalamus (MGN) confirming infection of AC. **d)** Reference section with the box

outlining the location of close up image shown in (E). **e)** Close up image of highlighted section in D showing axonal terminals in superficial L1 and L5/6. **f)** Schematic of the experiment verifying optogenetic excitation of AC cell bodies. **g)** Raw voltage trace from an example electrode in AC during AC photostimulation, verifying effective optogenetic recruitment of local neurons. 5 mW, 10 ms pulses @ 20 Hz. **h)** Schematic of the experiment to optogenetically stimulate AC cell bodies and record laminar LFP in V1. **i)** CSD and LFP profile in V1 during AC photostimulation (average of n=2 sessions in 2 animals). Note how pulsed AC stimulation gives rises to a repetitive CSD response (sink) in the superficial (< 150 μm) and middle/deeper layers (500-800 μm). Vertical dashed lines indicate repeated AC stimulation. **j)** Same as in (i), but for an example visual checkerboard stimulation for comparative purposes. **k)** The event-related potential (ERP) following photopulses (+5 to +20 ms after pulse) increases as a function of fiber power, suggesting optogenetic stimulation affects V1 LFP in a dose-dependent manner**.** The ERP response was obtained by averaging the absolute signal from channels over all cortical depths.



Extended Data Figure 4.9 – Auditory behavioral dominance in conflict trials is stable, independent of performance, and depends on relative stimulus timing. a) As a proxy for subjective saliency or arousal, we measured pupil dilation over time for saliency-matched auditory and visual trials. Cropped image shows pupil fit. Line and shading indicate mean ± SEM across N = 40 sessions from 9 mice. All data in this figure are from MST mice. **b)** Quantification of maximal pupil dilation. The effect of modality on pupil dilation was tested in a linear mixed model with fixed effects of hit/miss, saliency, modality and random effect of mouse ID. Whether it was a hit or miss had the largest effect ($F(1,7530)=1138.95$, $p=1.24*10^{-232}$), then saliency ($F(1,7524)=33.36$, $p=7.97*10^{-9}$, with no effect of modality ($F(1,7526)=1.53$, $p=0.2164$). This supports the idea that visual and auditory conditions were matched in subjective saliency. **c)** Auditory dominance in an example session. Raster plots show for each trial type licks and rewards at the auditory and visual lick spout (red and blue tick marks respectively) aligned to stimulus change (t=0). Colored zones indicate response window (0 to 1.5s). Gray: inter-trial interval. Licks before t=0 were spontaneous. Note how during conflict trials, auditory licks and rewards dominate. **d)** Dominance index (DI) heatmap (as in

Fig. 4.5b) for the only animal out of 17 MST mice) displaying visual dominance. **e)** A heatmap of the auditory dominance index for conditions binned based on performance (d-prime) on unimodal trials. This is in contrast to the analyses presented in the main text, where conditions were grouped based on the predetermined saliency gauged by psychophysical performance in previous sessions. The current analysis controls for changes in performance by reassigning each bin of the heatmap to d-prime levels within that session. It can be seen that performance-matched conflict trial conditions (along the diagonal) have positive dominance index values, confirming auditory dominance. **f)** The saliency-matched dominance index (smDI) for conflict trials that are matched in performance to unimodal trials (conditions along the bottom-left to top-right diagonal of *e*) is significantly different from zero (Wilcoxon signed rank test, n=17 mice, p=0.030). Grey dot is mean ± SEM, *p <0.05. In (f) to (l), each dot is the smDI of one animal. **g)** Auditory dominance was stable across the session, with auditory dominance computed on the first and second half of sessions being similar (Wilcoxon signed rank test, n=17 mice, p=0.492). **h)** Dominance was not correlated with visual performance (d-prime on unimodal trials of maximal visual saliency in the same sessions; r=-0.29, p=0.26). **i)** Dominance was not correlated with auditory performance (d-prime on unimodal trials of maximal auditory saliency in the same sessions; r=0.24, p=0.35). **j)** Dominance was not correlated with mean reaction time in visual trials (r=0.38, p=0.14). **k)** Dominance was not correlated with mean reaction time in auditory trials (r=0.09, p= 0.73). **l)** Reaction times on visual, auditory and conflict trials. For conflict trials, only saliency-matched conflicts are shown (Vsub + Asub, Vthr + Athr, etc.). Conflict trials were split based on choice. Mean ± SEM. **m)** We varied stimulus onset asynchrony between auditory and visual stimulus changes during conflict trials. The plot shows the percentage of auditory choice (red), visual choice (blue) or no lick (black) during saliency-matched threshold-level conflict trials as a function of stimulus onset asynchrony (SOA). A positive SOA value means that the visual change was presented first, followed by the auditory change. **n)** Purple error bars show mean and standard deviation of DI as a function of stimulus-onset asynchrony. Black line and gray shading show bootstrapped cumulative Gaussian fit of DI as a function of SOA (median and 95% confidence interval). Top error bar and dotted line indicate crossover point, i.e. fitted μ parameter (median and 95% confidence interval). Auditory dominance reverses once the visual stimulus change precedes the auditory change by 89.9 ms (95% CI: 47.7-138.7 ms). This is close to the difference in reaction time between saliency matched auditory and visual conditions: 110.5 ms on average. In other words, when the visual change preceded the auditory change by about 90 ms, auditory dominance was halfway to reversing into visual dominance. Further advancing the visual change in time completely reversed the dominance. This may reflect a scenario in which the visual evidence has instructed the decision-making system to an extent that subjects have already committed to a motor plan (namely to lick the visual spout) before the auditory evidence may take control. Similar temporal dominance of audition over vision has been reported in humans (Burr et al., 2009; Repp and Penel, 2002; Shams et al., 2000).

Chapter 5: Functional (ir)relevance of posterior parietal cortex during audiovisual change detection

Matthijs N. Oude Lohuis
P. Marchesi
C.M.A. Pennartz
U. Olcese

Published in:
Journal of Neuroscience (2022)
42(26):5229-5245
10.1523/JNEUROSCI.2150-21.2022

Abstract

The posterior parietal cortex (PPC) plays a key role in integrating sensory inputs from different modalities to support adaptive behavior. Neuronal activity in PPC reflects perceptual decision making across behavioral tasks, but the mechanistic involvement of PPC is unclear. In an audiovisual change detection task, we tested the hypothesis that PPC is required to arbitrate between the noisy inputs from the two different modalities and help decide in which modality a sensory change occurred. In trained male mice, we found extensive single-neuron and population-level encoding of task-relevant visual and auditory stimuli, trial history, as well as upcoming behavioral responses. However, despite these rich neural correlates, which would theoretically be sufficient to solve the task, optogenetic inactivation of PPC did not affect visual or auditory performance. Thus, in spite of neural correlates faithfully tracking sensory variables and predicting behavioral responses, PPC was not relevant for audiovisual change detection. This functional dissociation questions the role of sensory- and task-related activity in parietal associative circuits during audiovisual change detection. Furthermore, our results highlight the necessity to dissociate functional correlates from mechanistic involvement when exploring the neural basis of perception and behavior.

Significance Statement

The Posterior Parietal Cortex (PPC) is active during many daily tasks, but capturing its function has remained challenging. Specifically, it is proposed to function as an integration hub for multisensory inputs. Here, we tested the hypothesis that, rather than classical cue integration, mouse PPC is involved in the segregation and discrimination of sensory modalities. Surprisingly, even though neural activity tracked current and past sensory stimuli and reflected the ongoing decision-making process, optogenetic inactivation did not affect task performance. Thus, we show an apparent redundancy of sensory and task-related activity in mouse PPC. These results narrow down the function of parietal circuits, as well as direct the search for those neural dynamics that causally drive perceptual decision making.

Introduction

The construction of context-dependent representations of sensory inputs is required to inform adaptive decision making. This process involves multiple functions, such as processing of stimuli coming from different sensory modalities, evidence accumulation and integration with past information, and finally transformation of sensory information into an appropriate motor plan. The parietal cortex has been identified as a key hub for these functions through classical lesion studies in humans (Bender and Teuber, 1947; Denny-Brown et al., 1952; Holmes, 1918), including seminal studies on hemineglect (Kerkhoff, 2001; Vallar, 1998), and through extensive work in primates (Andersen and Cui, 2009; Bisley and Goldberg, 2010; Cui and Andersen, 2007; Freedman and Assad, 2006; Platt and Glimcher, 1999; Robinson and Goldberg, 1978). In rodents – a key animal model to study the circuit-level mechanisms of cognitive processes – the posterior parietal cortex (PPC) is located at the nexus of visual, auditory, and somatosensory cortices (Hovde et al., 2019; Wilber et al., 2015) and is bidirectionally connected to sensory and motor areas, as well as to other associative cortical areas (Oh et al., 2014; Shadi et al., 2020), homologous to primate PPC (Whitlock et al., 2008; Wilber et al., 2015). Neurons in rodent PPC respond to visual, auditory, and somatosensory inputs (Mohan et al., 2018; Nikbakht et al., 2018; Olcese et al., 2013; Raposo et al., 2014; Wallace et al., 2004). Sensory-evoked PPC responses have been described in naive animals (Olcese et al., 2013), but PPC is mostly recruited during task engagement (Pho et al., 2018). Neuronal activity in PPC has been shown to reflect aspects of perceptual decision making such as an animal's decision (Driscoll et al., 2017; Funamizu et al., 2016; Goard et al., 2016; Krumin et al., 2018; Pho et al., 2018; Runyan et al., 2017; Song et al., 2017), accumulated evidence (Hanks et al., 2015), trial history (Akrami et al., 2018; Hwang et al., 2017) and working memory (Harvey et al., 2012). This combination of sensory-, task- and choice-related activity suggests an important role in perceptual decision making, but under which conditions PPC is causally involved is less clear. Several studies find PPC necessary for making decisions based on visual stimuli (Driscoll et al., 2017; Goard et al., 2016; Harvey et al., 2012; Licata et al., 2017), but not for tactile (Guo et al., 2014a) or auditory information (Erlich et al., 2015; Licata et al., 2017). Other studies have suggested that PPC is causally involved in both visual and auditory tasks, but only when additional cognitive processes beyond simple sensorimotor associations are required (Akrami et al., 2018; Funamizu et al., 2016; Harvey et al., 2012; Hwang et al., 2017; Licata et al., 2017; Song et al., 2017; Zhong et al., 2019).

As PPC receives converging inputs from auditory and visual cortices and is strongly connected to (pre)motor areas, we tested the hypothesis that PPC is required to solve tasks in which signals from two different modalities have to be compared to reach a behavioral decision. We designed an audiovisual change detection task for which PPC is conjectured to arbitrate between noisy inputs from different sensory channels and help decide in which modality a sensory change occurred. In line with

our hypothesis, we found that PPC displayed rich visual, auditory, and decision-related activity that reflected sensory processing and predicted task performance. However, optogenetic inactivation of PPC did not affect either visual or auditory change detection. These results fundamentally question which function sensory- and task-related activity in PPC may fulfill, and to what extent the presence of rich sensory and task correlates reflects the role in perception and behavior of a cortical area.

Materials and Methods

Data and Code Availability

The data and code that support the findings of this study are available from the corresponding authors, M.O.L., C.M.A.P., and U.O., upon reasonable request.

Animals

All animal experiments were approved by the Dutch Commission for Animal Experiments and by the Animal Welfare Body of the University of Amsterdam. A total of 24 male mice was used from two transgenic mouse lines: PVcre (JAX 008069) and F1 offspring of Pvcre (JAX 008069) and Ai9-TdTomato cre-reporter mice (JAX 007909). Mice were at least 8 weeks of age at the start of experiments and group-housed under a reversed day-night schedule (lights were switched off at 8:00 and back on at 20:00). All experimental procedures were performed during the dark period.

Head bar implantation

Before the start of any experiment, mice were implanted with a custom-made titanium head-bar to allow head fixation. Mice were anesthetized with isoflurane and fixed in a stereotaxic apparatus. A circular head bar was positioned to include V1 and PPC bilaterally and glued and cemented to the exposed skull. Areas of interest were located based on stereotaxic coordinates (V1 relative to lambda: AP 0.0, ML ± 3.0, PPC relative to bregma: AP 1.9, ML ± 1.6 (Goard et al., 2016; Le Merre et al., 2018; Song et al., 2017)). Mice were allowed to recover for 2-7 days after implantation and were then habituated to handling and head-fixation before the start of the training procedure.

Audiovisual change detection task

Throughout experiments, mice were water-deprived and earned their daily ration of liquid by performing the behavioral task. Mice were head-fixed and two lick spouts were positioned symmetrically on the left and right side within reach of their tongue. Licks were detected by capacitance-based (training setups) or piezo-electric-based detectors (recording setup). Upon correct licking, 5-8 µl of liquid reward (Infant formula, Nutrilon) was delivered through the lick spout using gravitational force and solenoid pinch valves (Biochem Fluidics).

Stimuli were continuously presented throughout behavioral sessions. Visual stimuli were drifting square-wave gratings with a temporal frequency of 1.5 Hz and spatial frequency of 0.08 cpd at 70% contrast. Stimuli were presented with a 60 Hz refresh rate on an 18.5-inch monitor positioned at a straight angle with the body axis from the mouse at 21 cm from the eyes. In trials with a visual change the orientation of the drifting grating was instantaneously changed (e.g. from 150° to 180°) while preserving the phase. The auditory stimulus was a stationary Shepard tone (Shepard, 1964) composed of a center tone and multiple harmonics (2 lower and 2 higher harmonics). The center tones ranged a full octave spanning from 2^{13} Hz (=8372 Hz) to 2^{14} Hz (=16744 Hz). For each given Shepard tone in this stimulus set, the weight of the center and harmonic tones are taken from a fixed Gaussian weight distribution over all center and harmonic tones, in this case centered at $2^{13.5}$ (=11585 Hz). Stimuli were presented with a sampling rate of 192 kHz. Stimuli were high-pass filtered (Beyma F100, Crossover Frequency 5-7 kHz) and delivered through two bullet tweeters (300 Watt) directly below the screen. Sound pressure level was calibrated at the position of the mouse and volume was adjusted per mouse to the minimum volume that maximized performance (average ±70 dB). In trials with an auditory change, the stimulus was modified instantaneously from one Shepard tone to another with a different center frequency and associated harmonics, while preserving the phase across all compound tones. For example, an auditory change of ¼ octave would jump from 2^{13} to $2^{13.25}$. The amount of change determined stimulus difficulty (see also below).

Animals were trained to respond in a lateralized manner to sensory changes in each modality: lick to one side to report visual changes, to the other side for auditory changes (modality-side pairing was counterbalanced across mice). In other words, mice were required to simultaneously monitor both the auditory and visual modality and identify the sensory modality in which a change occurred.

Trials were separated by a random inter-trial interval (taken from an exponential distribution with a mean of six seconds, minimum three, and maximum twenty seconds). Trial types were pseudorandomly ordered by block-shuffling: every block of 10 trials contained a mixture of trial types in a fixed proportion but random order (8% catch trials=no change, 46% visual trials, 46% auditory trials). For instance, each block contained on average 46% of visual trials randomly interspersed among the other trial types. Directly after stimulus change, a response window of 1500 ms followed in which mice could obtain a reward by licking the correct side (no reward available in catch trials). The first lick on the correct side during the response window was immediately rewarded (median reaction time was 324 ms across all auditory hits and 407 ms across all visual hits).

It is not likely that mice solved the task using short-term memory or prolonged evidence accumulation, as mice could respond immediately and the observed reaction times are shorter than the time windows usually considered in behavioral tasks explicitly including evidence accumulation or working memory (~1000 ms)

(Akrami et al., 2018; Brunton et al., 2013; Odoemene et al., 2018). However, it can still be argued that this task could be solved for instance by keeping the previous orientation in memory and comparing it to the previous one.

For each trained animal, we presented five levels of auditory and visual amount of change that spanned the perceptual range. We fitted this psychophysical data (see below) to establish the perceptual threshold for the visual and auditory domains for each animal. To sample sufficient trials per condition in recording sessions, we used two levels of change (threshold and maximal). For noncontingently exposed (NE) mice, we used threshold values that matched those from trained animals.

As visual and auditory feature changes were associated with different motor actions (in trained but not in NE animals) a simultaneous auditory and visual change would present the animal with conflicting signals (the visual change predicts reward for licking left, auditory change predicts reward for licking right). In a subset of sessions, we introduced these conflict trials (25% of the trials – replacing unimodal trials) and registered the choice (side of the first lick). Both sides were rewarded in conflicting trials.

We compared multisensory trained animals (*MST*, n=17) with NE animals (n=7). For NE animals, the sensory environment was identical; both the auditory and visual stimuli were continuously presented with the same distribution of trial types and temporal statistics as for trained animals. However, these animals received rewards if they licked during a hidden response window that was temporally decorrelated from the stimuli. Spontaneous licks were therefore occasionally rewarded and this allowed us to compare intermittent licks, rewards, and stimuli between trained and noncontingently exposed mice. Each session was terminated after 20 trials of unresponsiveness and these last 20 trials were always discarded from all analyses.

We further excluded a subset of sessions from trained animals with poor behavioral performance due to (1) a high false alarm rate to auditory or visual spout (>50%, one session excluded), (2) a high lapse rate on easiest auditory or visual trials (a threshold of >30% hit rate on both visual and audio trials with a maximal amount of change had to be met, two sessions excluded for low auditory performance, nine sessions for low visual performance). Note that in this two-alternative *un*forced choice task, performance at chance level is not 50% with three response options (lick to visual spout, auditory spout, no lick) and depends on spontaneous lick rates and trial type distribution.

Viral injection

Once animals were trained to asymptotic performance, we aimed at optogenetically inactivating PPC by locally expressing Channelrhodopsin-2 in a cre-dependent manner in PVcre mice, therefore driving inhibitory interneurons. Mice were subcutaneously injected with the analgesic buprenorphine (0.025 mg/kg) and

maintained under isoflurane anesthesia (induction at 3%, maintenance at 1.5–2%) throughout the surgery. Small craniotomies (~100 µm) were made over the area of interest (V1 or PPC) using a dental drill. A glass micropipette backfilled with AAV2.1-EF1a-double floxed-hChR2(H134R)-EYFP-WPRE-HGHpA (titer: 7×10^{12} vg/mL, 20298-AAV1 Addgene) was slowly lowered in the cortex, and 25nL was injected at 700 µm and 400 µm below the dura each using a Nanoject pressure injection system (Drummond Scientific Company, USA). Further experiments continued after three weeks to allow for viral expression. A potential concern is that expression of ChR2 in deep PV-expressing pyramidal neurons (Tanahira et al., 2009) led to an increase rather than a decrease of activity in deep layers. We observed no such expression in deep pyramidal neurons of PPC (Figure 5.6B).

Neuronal recordings in PPC

We performed craniotomies on the day before starting extracellular recording sessions. Mice were anesthetized with isoflurane and small (~300-500µm) craniotomies over the areas of interest were made using a dental drill leaving the dura intact. Craniotomies were made in the left hemisphere over V1, PPC, as well as primary auditory cortex and medial prefrontal cortex. Only data from V1 and PPC was analyzed for this study. The data presented from these areas was partly recorded in the same animals. Data regarding optogenetic manipulations of V1 and PPC activity was gathered in different animals.

Extracellular recordings were performed on consecutive days with a maximum of 4 days to minimize damage to the cortical tissue. Microelectrode arrays of 32 or 64 channels (NeuroNexus, Ann Arbor, MI – A1x32-Poly2-10mm-50s-177, A1x64-Poly2-6mm-23s-160) were slowly inserted perpendicularly to the cortical surface until all recording sites were in contact with the tissue. To allow for tissue stabilization, the start of the behavioral task commenced at least 15 minutes after array insertion. For the recording session on the final day before perfusion, the array was covered in DiI (ThermoFisher Scientific) to facilitate post hoc visualization of the electrode tract. Neurophysiological signals were pre-amplified, bandpass filtered (0.1 Hz to 9 kHz), and acquired continuously at 32 kHz with a Digital Lynx 128 channel system (Neuralynx, Bozeman, MT).

Optogenetics

In sessions with optogenetic interventions, a random subset of trials (50% of trials) was associated with photostimulation. Photostimulation started at stimulus onset and continued until the animal made a choice. Two fiber-optic cannulas (ID 200 µm, NA 0.48, DORIC lenses) were connected to a 473 nm laser (Eksma Optics, DPSS 473nm H300) and positioned within 1 mm directly over the thinned skull at the area of interest. We performed extracellular recordings simultaneous with photostimulation in all mice to verify the effectiveness of inactivation and we adjusted the laser power for each animal to the minimum power that maximally inhibited neural activity. The horizontal offset between the fiber tip and insertion

site of the microelectrode array was minimized within the limited space constraints and measured approximately 200-400 μm (for comparison: PPC measures approximately 1.0-1.5 by 1.0-1.5 mm, depending on delineation (Lyamzin and Benucci, 2019)). The range of laser powers used was the same for PPC animals compared to V1 animals (2-15 mW at the tip of each fiber-optic cannula, corresponding to 15.9 – 119.3 mW mm^{-2}). This laser power in combination with our optogenetic approach was previously shown to lead to effective spatial inhibition across our infected target area (Li et al., 2019). To allow rapid control over light delivery, laser beam continuity was controlled by a shutter (Vincent Associates LS6 Uniblitz). We stimulated with 10 ms pulses at 20 Hz (40 ms off, 20% duty cycle). For the photoinactivation of V1, a stimulation scheme was used in which the pulse and interpulse duration were variable with an average of 20 Hz and 75% duty cycle. The higher duty cycle of V1 inhibition versus PPC inhibition (75% versus 20%) is unlikely to explain the difference in effect on task performance for two reasons. First, when we investigated spiking activity relative to single laser pulses we observed no rebound activity during the interpulse interval (see Figure 5.6D). Second, the same stimulation protocol has been used in the same lab to effectively silence higher visual areas to study the impact on V1 (Chapter 2). To prevent light from reaching the eye of the mouse, the fiber-optic cannulas were sealed with black tape leaving only the tip exposed. Furthermore, animals performed the task in an environment with ambient blue pulsating light.

Pupil monitoring

The left eye was illuminated with an off-axis infrared light source (IR-LEDs, 850 nm) positioned to yield high contrast illumination of both the eye and whisker pad. A near-infrared monochrome camera (CV-A50 IR, JAI) coupled with a zoom lens (Navitar 50 mm F/2.8 2/3" 10MP) was positioned at approximately 30 centimeters from the mouse to capture a view of the lick spouts and face of the mouse. A frame-grabber acquired images of 752x582 pixels at 25 frames per second. The pupil size and position were extracted from the obtained videos by labeling the pupil center and 6 radially symmetric points on the edge of the pupil using DeepLabCut (Mathis et al., 2018) and pupil size was quantified as the surface area of an ellipse fit to these points.

Histology

At the end of the experiment, mice were overdosed with pentobarbital and perfused (4% paraformaldehyde in phosphate-buffered saline). The brains were recovered for histology to verify viral expression and recording sites. We cut coronal and flattened cortical sections as described previously (Lauer et al., 2018). For coronal sections, area borders were drawn by aligning and overlaying the reference section from the atlas (Paxinos and Franklin, 2004). For flattened cortical sections areas were identified based on cell densities aligned to reference maps (Gămănuţ et al., 2018).

Data analysis

Unless otherwise stated, all data were analyzed using custom-made software written in MATLAB (The MathWorks, Natick, MA).

Behavior: analysis of performance

Behavioral hit rates were fit with a multi-alternative signal detection model (Sridharan et al., 2014). This model extends signal detection theory (Green and Swets, 1966) for multiple signals and has been designed to accurately and parsimoniously account for observer behavior in a detection task with multiple signals. In our behavioral task, these are the auditory and visual signals to be reported at different lick spouts. In this model, the decision is based on a bivariate decision variable whose components encode sensory evidence in each modality, and decision space is partitioned in three regions (miss: neither evidence is strong enough, auditory response and visual response). In a given trial, the observer either chooses to report nothing (no licking) or report visual or auditory stimuli (by licking left or right) if the decision variable exceeds a particular cutoff value, the "criterion" for each signal (the animal's internal signal threshold for responding, in terms of signal detection framework). We fit two versions of this model. In sessions with two levels of change per modality (threshold and maximum), we fit the d-prime (d') and criterion (c) to the behavioral response rates for each stimulus difficulty separately. This consists of fitting four free parameters: the d-prime parameters (d'_{vis}, d'_{aud}) and criterion parameters (c_{vis}, c_{aud}). In sessions with four levels of change per modality, we fit the behavioral response rates by fitting a fixed criterion and a d-prime (d') at each stimulus difficulty that was described by psychophysical function (three-parameter hyperbolic function). The d-prime at each stimulus difficulty follows from:

$$d'_i = d'_{max} \cdot x_i^n / (x_i^n + s_{50}^n) \qquad (Eq.\ 5.1)$$

where d'_{max} is the asymptotic d', s_{50} is the stimulus strength at 50% of the asymptotic value, n is the slope of the psychometric function and x_i is the amount of change. This consisted of fitting a total of 8 free parameters: d'_{max}, n, s_{50}, and c for each modality. These eight parameters are presented in Figure 5.8.

For a detailed description of how the d-prime and criterion subsequently relate to response rates, we refer the reader to (Sridharan et al., 2014). This analysis was used before the start of experimental sessions to establish for each animal the perceptual threshold and therefore auditory and visual stimulus difficulty to use, as well as quantify behavioral performance on control and optogenetic trials.

Behavior: regression model of choice

To test whether PPC inactivation affected other factors beyond performance averaged over trial types we constructed a regression model of behavioral choice (Akrami et al., 2018; Busse et al., 2011; Hwang et al., 2017). We selected a

multinomial logistic regression model, as in our behavioral task the animal was presented with three discrete choice options (lick to visual spout, lick to auditory spout, or no-go). In this model, each regressor has two weights, where positive weights increase the probability of responding relative to a third reference option. We chose 'no-go' (not licking) as the reference option, such that the linear sum of weights across regressors determine the probabilities of the choice on a given trial through:

$$\log\left(\frac{p_a}{p_{n\ -g}}\right) = b_a \quad + \beta_a \quad \square X \tag{5.2}$$

$$\log\left(\frac{p_v}{p_{n\ -g}}\right) = b_v \quad + \beta_v \quad \square X \tag{5.3}$$

Where b is a bias parameter, β are sets of regression coefficients (or weights), and X the matrix with the values of regressor variables for every trial. As regressors, we used information from the current and the previous trial. Only information from up to one trial in the past was used, as this captured the strongest effects on behavioral choice and allowed us to test the effects of PPC inactivation. First, to construct a null model we included a random predictor with random values between 0 and 1 (abbreviated to *N* in Figure 5.9). This null model can already predict behavioral choice above chance by matching choice fractions. Second, we included the within-session trial number to account for non-stationarity in behavioral choice due to satiation (*T)*. Third, for sensory information we used the amount of visual and auditory change (degree of grating orientation and pitch change, respectively) as a scalar variable (both log-transformed to account for logarithmic sensitivity in sensory systems) (S_v, S_A), and sensory stimuli on the previous trial (S_{V-1}, S_{A-1}). Fourth, reward on the previous trial was captured in two binary regressors per modality (R_{V-1}, R_{A-1}; 0=not-rewarded, 1=rewarded). Fourth, we included choice history (C_{-1}). Because this was a detection task with sensory stimuli at or near the perceptual threshold, licks outside 'trials' to the auditory or visual lick spout could be interpreted as reports of perception. Therefore, we included for choice history a binary predictor that reflected the side of the last lick (0= last auditory lick, 1= last visual lick). Lastly, to capture the effect of PPC photoinactivation on choice a binary predictor of photostimulation was included (*O*).

All sessions of single animals were concatenated as if it were one session (N=7.2 sessions on average). The model was fit on this concatenated trial data per animal (N=17 animals, M=58.652 trials in total) using the *glmnet* package in Matlab (Friedman et al., 2010) with elastic-net regularization (alpha=0.95) and 3-fold cross-validation. Regularization parameter *lambda* was maximized while not decreasing cross-validated performance. Model fit quality was assessed as the fraction of held-out test trials in which the estimated choice (choice with the

highest probability) matched the actual choice. We term this cross-validated model performance and it is shown in the figure. In partial models, we used a subset of variables and performed the same regression procedure; for instance a model with and without including photostimulation as a predictor to assess the effect of PPC inactivation on model performance. The analysis of the effects of photoinactivation was done on the five animals with PPC inactivation.

Neural data processing

Before spike sorting, the median of the raw trace of nearby channels (within 400 μm) was subtracted to remove common noise artifacts. Automated spike sorting and manual curation were done using Klusta and the Phy GUI, respectively (Rossant et al., 2016). During manual curation, each putative single unit was inspected based on its waveform, autocorrelation function, and its firing pattern across channels and time. High-quality single units were included as having (1) an isolation distance higher than 10 (Schmitzer-Torbert et al., 2005); (2) less than 0.1% of their spikes within the refractory period of 1.5 ms (Bos et al., 2017; Vinck et al., 2015b), (3) stable presence throughout the session. This latter was quantified by binning the firing across the entire session (~50 minutes) in 100 time bins and only including neurons that spiked in more than 90 time bins. We recorded a total of 671 neurons from 14 animals over 32 sessions that met our criteria.

For Figure 5.1D and 5.1F, spike times were binned in 1 ms bins, convolved with a Gaussian window with 50 ms standard deviation, and z-scored by subtracting the mean baseline activity and dividing by the standard deviation of all baseline periods (-1 to -0.2 seconds before stimulus). For the single neuron encoding model, spikes were binned in 25 ms bins and convolved with a Gaussian window with a 50 ms standard deviation.

Single neuron encoding model

We constructed an encoding model that allowed us to model, for single neurons, the time-dependent effects of all measured variables related to the task and the animal's behavior simultaneously on single-trial neuronal activity. This approach is particularly useful to disentangle the different events that contribute to heterogeneous responses in associative regions such as parietal cortex (Park et al., 2014).

We included six categories of predictors: visual stimuli, auditory stimuli, reward, licking movement, pupil size, and trial history. Binary variables (stimulus present or not, last trial rewarded or not; all variables except for pupil size were binary) were modeled with a series of temporal basis functions (raised cosines) that spanned the relevant epoch of influence to fit time-dependent modulation of neuronal responses by these predictors.

For the auditory and visual predictors, we used two kernels with 100 ms standard deviation that spanned the first 200 ms post-stimulus to capture the early spiking activity and 10 kernels with 200 ms standard deviation that spanned from 0 to 2000 ms post-stimulus to capture the late, sustained response. A separate predictor set was used per combination of orientation/frequency x amount of change. For reward variables, we used 10 kernels with 200 ms standard deviation that spanned from 0 to 2000 ms relative to stimulus change in hit trials (visual hit, audio hit) and 10 predictors that spanned -500 ms to +1500 ms relative to reward. We used *reward* as a single term to refer to hit-trial specific activity (rewarded visual and auditory hits). This definition was used also in tensor component analysis and population decoding (see below). Note that this definition captures activity related to correct detection and report as well as reward-related activity and this encoding model dissociated this from other confounds (such as licking and arousal) by exploiting the trial-by-trial variability in timing. For licking movement, we used three kernels that spanned -200 to +400ms relative to each lick, split by lick side. To capture arousal effects, the z-scored pupil area was included in the predictor set: with original timing and two temporal offsets (-800 ms and -400 ms). We included three binary history-dependent predictors, capturing the modality of the previous trial (visual or auditory), reward (hits or not), and choice (lick to visual or auditory spout). Lastly, the trial number was included to account for non-stationarity in firing rate across the entire session due to, for example, motivational signals and electrode drift, but was not reported in figures.

The number, width, and spacing of temporal basis functions were selected by optimizing the variance explained on a diverse subset of representative neurons (Runyan et al., 2017). For example, the use of trial-spanning history predictors was based on the fact that history had an offset effect throughout the trial duration in example neurons (Figure 5.3). Explained variance both on trial averages and single trials was comparable to previously reported studies (Runyan et al., 2017; Steinmetz et al., 2019).

This resulted in a predictor matrix of size P x T for each neuron, where P is the number of predictors and T is the number of total time bins. The encoding model was fit on concatenated single trials and T is therefore the number of trials (typically 200-500 trials) multiplied by the number of time bins per trial (100 time bins; -0.5 to +2 sec relative to stimulus change, 25 ms time bins). The encoding model (a generalized linear model, GLM) was fit with a Poisson link function to single neuron activity with the *glmnet* package in Matlab (Friedman et al., 2010). We used elastic-net regularization ($\alpha = 0.95$) and 5-fold cross-validation. To maximally punish weights without losing model fit quality, lambda was maximized while keeping the cross-validated error within one standard error of the minimum. We quantified model performance by assessing the 5-fold cross-validated explained variance (EV) in two ways. First, we computed EV over all concatenated firing rate bins (over all single trials). Second, we computed EV on the concatenated

firing rate bins of the average firing rate for four main trial type conditions with most trial counts (visual and auditory hits and misses) (Musall et al., 2019; Runyan et al., 2017).

To quantify the contribution of subsets of predictors we calculated the single-trial EV for a firing rate prediction based on only those predictors. In other words, how much of the firing rate variance is explained by only considering the weights from, for example, visual variables. This value was compared to a shuffled distribution where the EV was computed using the predicted firing rate and the actual firing rate from shuffled trials (N=1000 shuffles). A neuron was deemed to significantly encode this variable if it exceeded the 99% percentile of this shuffled distribution. This permutation test approximately labeled variable encoding in single neurons if at least 1% of the single-trial variance was explained, as we found very similar results when we simply thresholded on 1% EV.

To quantify whether the joint encoding of variables was significantly different from a random distribution across neurons, we shuffled (N=1000) the vectors of neuron indices that significantly encoded visual, auditory, and reward variables relative to each other and recomputed joint or unique encoding, i.e. recomputing the Venn diagram. For instance, with 50% of neurons encoding visual and 50% encoding auditory variables, this shuffling procedure would generate percentages of joint audiovisual encoding neurons around 25%, against which the actual percentage was tested (exceeding the 2.5% or 97.5% percentile, corresponding to a two-sided test with $p<0.05$).

Population decoding analysis

We tested whether PPC ensembles were responsive to audiovisual stimuli and reward by training decoders to classify (1) audio vs. catch trials, (2) visual vs. catch trials, (3) visual vs. audio trials, and (4) rewarded vs. non-rewarded trials. In (1), (2), and (3) we included only trials with large stimulus changes. Decoding was performed on recordings that contained at least 15 neurons and 15 trials per class. Spikes were binned using a sliding window of 100ms with 50ms increments, whereas for the insets, showing broader temporal dynamics, we used a window of 500ms moved with 250ms increments. Trials were aligned to the moment of stimulus change (Figure 5.4A-H) or the timing of the first response lick (i.e. the first lick occurring at least 100ms after stimulus change, Figure 5.4I,J). When trials were aligned to stimulus change, temporal bins containing data before and after stimulus change (t=0) were excluded from the analysis. Decoding was performed using a random forest classifier with 200 trees, as implemented in Scikit-learn (Pedregosa et al., 2011), and we employed a 5x5 cross-validation routine with stratified folds (which preserves the proportions of samples of the two classes in each fold). The average accuracy obtained in the cross-validation routine was corrected by subtracting the average accuracy on 300 surrogate datasets in which the trial labels were randomly permuted to obtain the improvement in decoding

accuracy beyond chance level. For every time point, we tested whether the corrected accuracy was significantly different from 0 using Wilcoxon signed-rank test; p-values were corrected for multiple comparisons using the Benjamini–Yekutieli method, with a significance level of 0.05. We additionally trained decoders to classify the orientation of the drifting gratings and the frequency of the audio tone (Figure 5.4G,H). Before decoding visual and auditory stimulus identity, we grouped the two pairs of orientations/frequencies, to obtain a two-class classification problem, and included only trials in which the stimulus changed from one pair of stimuli to the other (max changes).

Tensor Component Analysis

We applied tensor component analysis, TCA, (Williams et al., 2018) to recordings that contained at least 15 neurons. As the objective and similarity plot did not demarcate a fixed number of components to decompose our data, we chose a rank of 8 and tested the robustness of our analysis to different choices of rank (5 and 10 components, data not shown). As a way of identifying components representing noise and drift in the recordings, we filtered out components that were expressed in less than 20% of units or less than 20% of trials. For each component, we measured with a ROC-AUC score how well the trial factors discriminated (1) audio trials vs. catch trials, (2) visual trials vs. catch trials, and (3) correct trials vs. incorrect trials. For each AUC score, we computed an associated p-value by generating a null distribution of AUC values calculated on randomly permuted trial labels and computing the fraction of values in the null distribution which were equal or larger than the observed AUC score. AUC scores were considered significant with $p<0.05$. For Figure 5B,C components were labeled based on the contrast associated with the highest AUC score and included if the highest AUC score had an associated $p<0.05$.

Statistics

We used Bayesian statistics throughout the manuscript (Jeffreys, 1939; Rouder et al., 2009) to facilitate intuitive interpretation of the strength of evidence as well as establish evidence of the absence of effects (Keysers et al., 2020). Bayesian statistics assess the likelihood of the data under both the null and the alternative hypotheses. In most cases, we report the Bayes Factor that corresponds to the ratio of likelihoods $p(data|H_1)/p(data|H_0)$, abbreviated to BF. For instance, BF=10 would mean that the data are 10 times more likely under H1 than H0 providing very strong support for H_1, while BF=0.1 would mean that the data are 10 times more likely under H_0 than H_1 providing very strong support for H_0. Generally, a BF between 1/3 and 3 indicates that the data is similarly likely under H_1 and H_0 and that the data thus does not adjudicate which is more likely. A BF below 1/3 or above 3 is interpreted as supporting H_0 or H_1, respectively, corresponding roughly to $P<0.05$ for moderate sample sizes (Jeffreys, 1939). Evidence of absence of an effect, where the null hypothesis is more likely given the data, is denoted by the

hashtag symbol ('#') in figures, next to the standard asterisk symbol ('*') for evidence for the alternative hypothesis.

We also performed classical frequentist statistics (i.e., calculating the probability of observing the data given a hypothesis) for each test and found nearly identical results, except for one statistical test in Figure 5.8C, where the auditory threshold was slightly affected by PPC inactivation.

We made four exceptions and used the classical frequentist test when we required a significance threshold without an available Bayesian alternative: (1) chi-square test for fractions for significant increased and decreased fractions of PPC neurons (Figure 5.1D) and (2) AUC values per TCA component versus a shuffled distribution (Figure 5), (3) significant time points of decoding performance (Figure 5.4), (4) significant (joint) encoding of stimulus and behavioral variables versus shuffled distributions (Figure 5.3).

Results

We trained mice to continuously and simultaneously monitor audiovisual stimuli and respond to one lick spout for changes in auditory frequency and the other spout for changes in the orientation of a drifting grating (Figure 5.1A). In other words, mice were required to identify the sensory modality in which a change occurred and could respond as soon as a change was detected. Behavioral performance was analyzed using a multi-alternative signal detection model (Sridharan et al., 2014). This model extends classical signal detection theory to distinguish perceptual sensitivity from choice bias in detection tasks with multiple signals (in this case visual and auditory changes) – see Methods. Trained animals selectively reported both orientation and frequency changes and performance increased as a function of the amount of change (Figure 5.1A). Animals were mostly tested in these unisensory trials, in which either a visual or an auditory change were presented, but never simultaneously. We also presented some truly multisensory trials, which will be discussed later. Mice lateralized their licks well, with occasional licks to the incorrect spout (10.5% errors at the visual lick spout, 14.9% errors at the auditory lick spout).

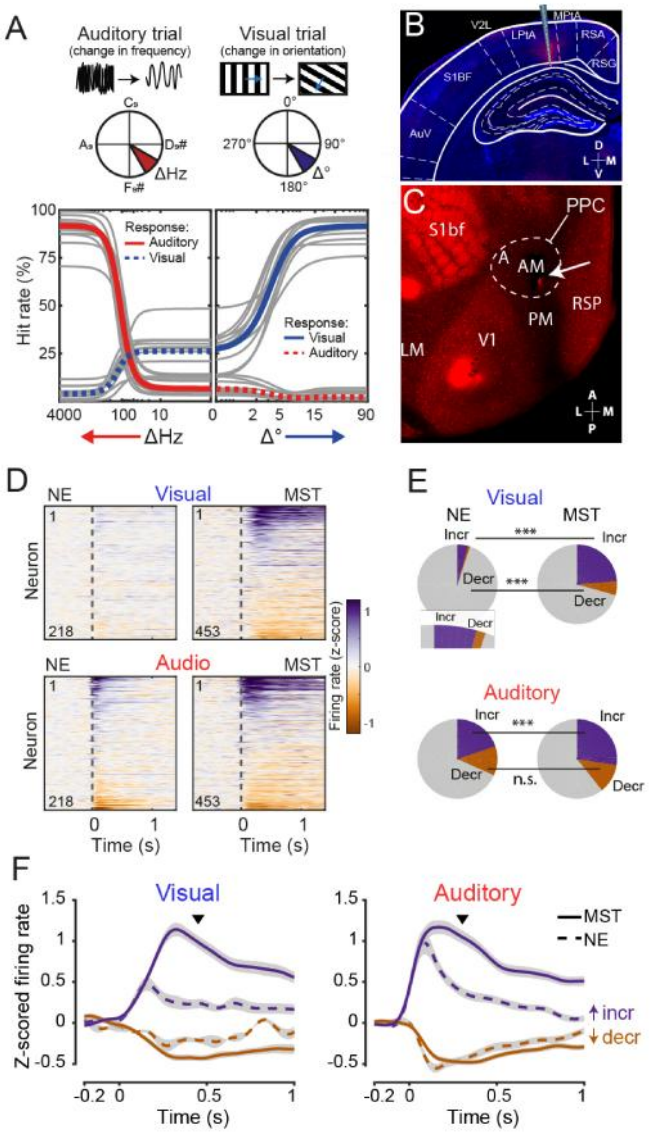
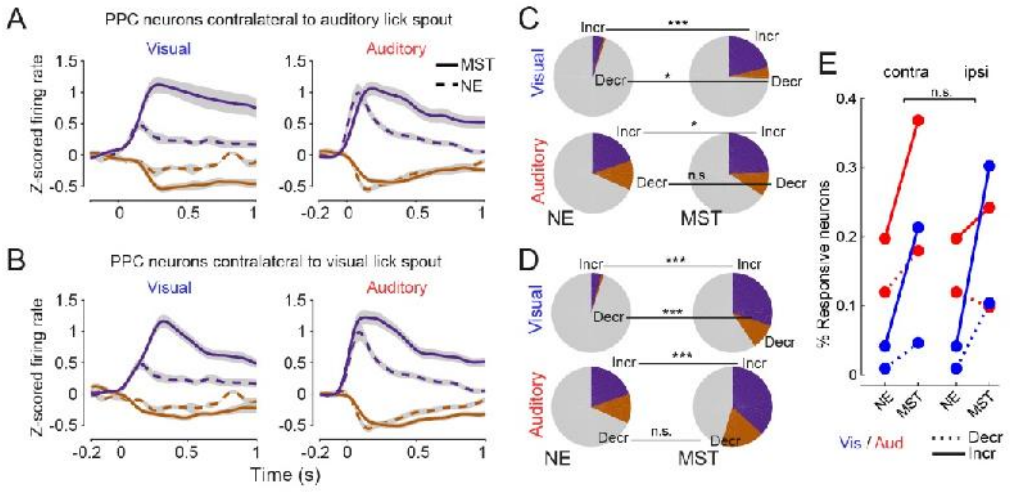


Figure 5.1: Posterior parietal cortex is recruited during audiovisual change detection. (A) *An auditory stimulus (Shepard tone) and a visual stimulus (drifting grating) were continuously presented. Headfixed mice had to report frequency changes (e.g. lick left on auditory trials) and orientation changes (lick right on visual trials). Gray lines show the psychometric fit of behavioral hit rates for individual sessions for an example mouse and colored lines the average across sessions (n=7 sessions). Dotted lines are licks to the incorrect spout. Licks at zero orientation and frequency change are false alarms (center of panels). The fraction of trials without a licking response is not shown but can be inferred from the figure as they sum to 100% together with shown response rates.* ***(B)*** *We recorded single-unit activity in PPC. Image shows histological verification with DAPI-staining in blue and electrode tract stained with DiI (in red) overlaid with reference section from (Paxinos and Franklin, 2004). LTPa: Lateral Parietal association area, V2L: lateral secondary visual cortex. RSA/G: retrosplenial cortex (a)granular part.* ***(C)*** *Flattened cortical section in a PvCre-tdTomato mouse providing landmarks through cell densities in red (Gămănuţ et al., 2018), with additional DiI-stained electrode tracts in V1 and PPC (also red): V1: primary visual cortex, LM: lateromedial, PM: posteromedial, A: anterior, AM: anteromedial, RSP: retrosplenial, S1 bf: barrel field of primary somatosensory cortex.* ***(D)*** *Heatmap of trial-averaged z-scored activity over all recorded neurons aligned to stimulus change. Each row corresponds to one PPC neuron recorded in noncontingently exposed animals (left panels; n=218 NE neurons) and multisensory trained animals (right panels; n=453 MST neurons), sorted by post-stimulus activity.* ***(E)*** *Fraction of neurons with significantly increased or decreased firing rate (Bayesian paired one-sided t-test BF>3, pre-stimulus -500 to 0 ms, post-stimulus 0 to 500 ms; BF corresponds to the ratio of likelihoods of the alternative and null hypothesis) is increased in MST versus NE animals (Chi-square two-sample tests: visual-increase, $p<0.001$, visual-decreased, $p=0.003$, auditory-increase, $p=0.031$, but not for auditory-decrease, $p=0.998$). Inset in top left panel is a close-up of the visually responding fractions in NE mice.* ***(F)*** *Z-scored firing rate for*

neurons with increased and decreased rate for visual and auditory trials separately show increased amplitude of modulation for trained versus naive mice. Black tick mark indicates median reaction time.

After establishing psychophysical performance, we only used two levels of change for each modality for subsequent sessions: threshold (individually titrated per animal) and maximal. We first wondered what types of task-relevant activity the PPC displays. We compared this cohort of multisensory-trained mice (MST) to another cohort of animals that was noncontingently exposed to the stimuli (NE). NE mice were not trained to detect changes but were pseudorandomly rewarded for spontaneous licks. We recorded single-unit activity in mouse PPC (Figure 5.1B,C), which corresponds to cortical areas also referred to as mouse Lateral Parietal Association cortex (LTPa; Paxinos and Franklin, 2004) or Anteromedial visual cortex (AM; Driscoll et al., 2017; Wang and Burkhalter, 2007). We found that the fraction of neurons responding to visual stimuli with changes in firing rate significantly increased from 5% in naive animals to 29.3% in trained animals, and from 31.6% to 39.3% for auditory stimuli (Figure 5.1D,E). In contrast to what is commonly reported in primary sensory areas, sensory stimuli both increased and decreased firing activity of PPC neurons (Figure 5.1F). The coupling between modality and rewarded lick spout (left or right) was counterbalanced across mice. We found, however, no systematic relationship for both modalities between the associated lick spout and evoked responses in contralateral PPC (Figure 5.2). We therefore show that, in addition to visual stimuli (Pho et al., 2018), also auditory stimuli evoke activity increments and decrements in PPC during task performance compared to passive stimulation.

(figure on the previous page) Figure 5.2: Increased auditory and visual responsiveness irrespective of modality-side pairing. *The modality-reward side pairing (which lick spout was associated with reporting changes in which modality) was counter-balanced across MST mice, while PPC was*



always recorded in the left hemisphere. To investigate whether the modality-reward side pairing was associated with different dynamics (e.g. increased auditory responsiveness in PPC if the auditory lick spout was contralateral), we repeated Figure 1E,F for neurons based on the pairing of the animals they were recorded from. Data from NE mice was the same as modalities were not

associated with specific lick spouts. Results were qualitatively similar for both sets of neurons. ***A)*** *Same as Figure 1F, but for PPC neurons contralateral to auditory lick spout.* ***B)*** *Same as Figure 1F, but for PPC neurons contralateral to visual lick spout.* ***C)*** *Same as Figure 1E, but for PPC neurons contralateral to auditory lick spout. Chi-square two-sample tests: visual-increase, p<0.001, visual-decreased, p=0.0149, auditory-increase, p=0.018, but not for auditory-decrease, p=0.424.* ***D)*** *Same as Figure 1E, but for PPC neurons contralateral to visual lick spout. Chi-square two-sample tests: visual-increase, p<0.001, visual-decreased, p<0.001, auditory-increase, p<0.001, but not for auditory-decrease, p=0.143).* ***E)*** *We tested directly whether more neurons were modulated during trials associated with a contralateral response (left PPC neurons during trials rewarded at the right lick spout) versus ipsilateral trials. The rise in the fraction of significantly responsive neurons was not significantly different based on laterality (n=8 fractions, BF=0.35).*

Sensory and behavioral correlates are heterogeneously distributed across single parietal neurons

We next focused on characterizing single-neuron responses in relation to stimuli and behavioral decisions. In trained animals, these were highly heterogeneous (Figure 5.3A-C). We found neurons that responded specifically to visual (example 1), or auditory stimuli (example 2), or were bimodally responsive (examples 3, 4). Additionally, the firing rate of many neurons was modulated during rewarded, but not unrewarded trials (examples 5, 6) or showed a mix of stimulus and reward-related activity (example 7; '*reward*' refers specifically to hit trials and thus captures activity related to stimulus detection, report, and reward, see Materials and Methods). Firing rates were further modulated by spontaneous licks outside trials (example 8), as well as by the previous trial's outcome (reward or not) and choice (left or right lick), indicating a history effect (examples 9, 10).

To capture and quantify the encoding of such heterogeneous sensory and task-related variables in the firing rate of single PPC neurons, we implemented a kernel-based regression model (Park et al., 2014; Runyan et al., 2017). This model accurately predicted single neuron firing rates (dashed lines in Figure 5.3A), explaining on average 46.1% of trial-averaged and 11.6% of trial-to-trial firing rate variance during 0-500 ms post-stimulus. This trial-to-trial variance was jointly explained by visual (12.1%) and auditory (23.4%) stimulus features, as well as

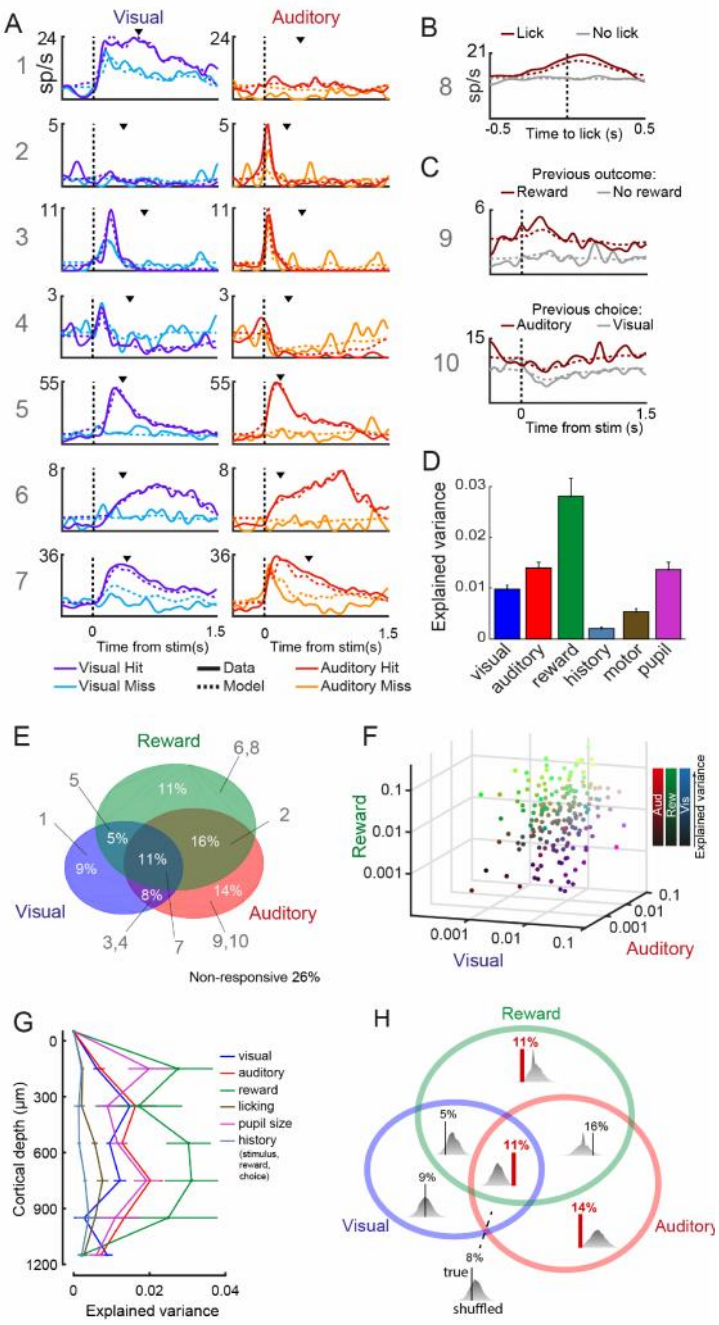


Figure 5.3: Heterogeneous task-related responses in single parietal neurons. ***(A)*** *Peri-stimulus time histograms (PSTH) for 7 example neurons for visual and auditory trials split by decision (hit or miss; licks to incorrect spout omitted). The dashed line shows the predicted firing rate based on a kernel-based regression model. Neurons responded to visual (1) and auditory changes (2), or changes in both modalities (3,4) and were modulated by decision (1,5,6,7). Black tick marks indicate mean reaction time in that session.* ***(B)*** *PSTHs for example neuron (8) aligned to spontaneous licks during inter-trial intervals (or random non-licking time points).* ***(C)*** *PSTHs for two other example neurons conditioned on the previous trial: reward (example 9), and choice (10), indicating history effects. We found no clear examples of neurons encoding stimulus history.* ***(D)*** *Explained variance of the trial-to-trial firing rate across 0-500 ms post-stimulus (this window includes lick response and reward delivery) for each category of predictors (Bayesian Anova, main effect of predictor group, $BF=7.9 \times 10^{13}$). Mean ± SEM.* ***(E)*** *Venn diagram showing the percentage of all PPC neurons that significantly encoded visual, auditory or reward, but also all combinations thereof (tested against random trial permutations, $p<0.01$, corresponding roughly to >1% EV). Gray numbers indicate where example neurons in A-C are located.* ***(F)*** *Scatter plot of explained variance by vision, audition, or reward. Each dot is a single neuron, with the RGB color scaled by the amount of EV for each variable (light=high EV; dark = low EV). Logarithmic axes were used to capture the spread in EV. The absence of clear clustering indicates heterogeneous mixed selectivity across the PPC population.* ***(G)*** *Average variance explained by each predictor for neurons based on cortical depth, showing that explained variance was not due to a subset of neurons localized to specific layers. Lines show mean ± SEM. N=483 neurons.* ***(H)*** *We tested*

whether the distribution of visual, auditory, and reward encoding across neurons was different from a random distribution based on chance (given the number of neurons significantly encoding each of these variables). Each inset in the Venn diagram shows the shuffled distribution and the actual percentage (in red if exceeding the 1% or 99% percentile of the shuffle distribution). The fractions of neurons exclusively encoding auditory stimulus features (14%) or reward (11%) were significantly lower than shuffled, while the percentage of neurons encoding all three variables (11%) was higher than that expected based on a random distribution of encoding across PPC neurons. So, joint encoding of auditory stimuli, visual stimuli, and reward was more common than exclusive coding of auditory stimulus and reward, and also more common than what would have been expected from a random distribution of encoding variables across neurons.

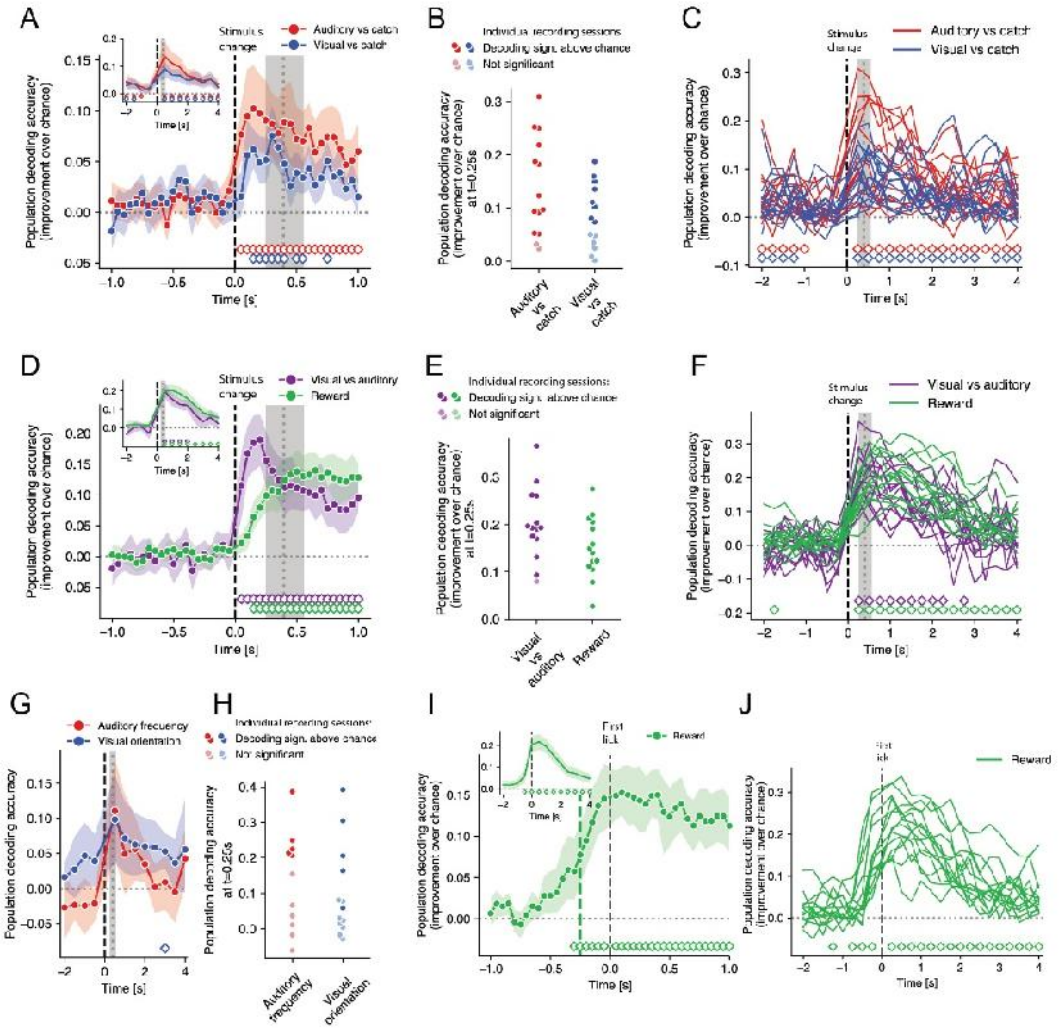
reward (37.6%), licking movement (7.8%), pupil size (16.3%), and trial history (stimulus, choice, and reward on the previous trial; 2.9%) (Figure 5.3D). We investigated whether these variables were encoded by distinct or the same neurons and focused on the variables most relevant to trial-by-trial performance (vision, audition, and reward). We found a mixed profile, with some neurons coding a single variable, while other neurons encoded two or all three of these variables simultaneously (Figure 5.3E), in line with the example neurons shown in Figure 5.3A. Visual, auditory, and reward coding were smoothly distributed across the neuronal population, without any clear clustering or organization (Figure 5.3F,G). Moreover, the number of neurons encoding all three variables was higher than expected based on chance (Figure 5.3H). Thus, single neuron correlates of audiovisual change detection were heterogeneously distributed, and not segregated, across the neuronal population.

Structure and content of parietal ensemble activity

To further investigate how information is globally encoded in PPC, we focused on how parietal population activity coded task-relevant variables relative to stimulus onset and decision. We used population decoding on ensemble activity recorded in individual experimental sessions and found that PPC activity sufficed to decode the presence of a sensory change (Figure 5.4A-C; significant decoding in 12/14 sessions for audition, 8/14 for vision), as well as the sensory modality of this change (Figure 5.4D-F; significant in 13/14 sessions for sensory modality). This suggests parietal activity does not indiscriminately respond to any sensory change, but that different subspaces encode visual vs. auditory information (Raposo et al., 2014). It was even possible to decode the sensory feature (specific orientation or frequency) in a subset of sessions (Figure 5.4G,H; significant in 5/13 sessions orientation, and 5/11 for tone frequency). Furthermore, population activity sufficed to decode the animal's decision (Figure 5.4F,I,J, significant in 14/14

sessions for reward), already about 250 ms before reaction time (dotted line in Figure 5.4I).

Figure 5.4: Parietal ensemble activity suffices to decode stimulus and decision. ***(A)*** *Accuracy of population decoding relative to chance plotted over time for random forest classifiers trained to discriminate visual versus catch trials (blue) and auditory versus catch trials (red). Trials were*



aligned to the time of stimulus change. Accuracy above chance was obtained by subtracting from the decoding accuracy the average decoding accuracy obtained after repeatedly shuffling the trial labels (for absolute decoding accuracies see Table 1). Diamonds indicate time points at which the distribution of decoding accuracy across recordings was significantly different from chance (Wilcoxon signed-rank test, p<0.01). Shaded bands indicate 95% confidence intervals. The vertical grey dotted line and shaded band mark the median and interquartile range of lick times, respectively. Inset shows the same decoding over a longer time window with larger temporal bins (similar in D and I). ***(B)*** *Decoding of auditory versus catch and visual versus catch at t=0.25s (time bin extending from 0 to 0.5s). Each dot is a session; where the p-value for each session was computed as the fraction of decoding accuracies obtained on shuffled data which were larger than*

the observed accuracy (similar in E and H). ***(C)*** *Same as A, but for individual recording sessions and longer time windows. Average decoding performance was not driven by a subset of sessions but was possible from nearly all recorded PPC populations.* ***(D)*** *Same as A, but for sensory modality and reward, i.e. discriminating visual versus auditory trials (purple), and rewarded versus unrewarded trials (green).* ***(E)*** *Same as B, but for sensory modality (visual versus auditory trials) and reward (rewarded versus unrewarded trials). Decoding performance was significantly above chance ($p<0.05$) in 13 out of 14 sessions for sensory modality, and 14 out of 14 for reward.* ***(F)*** *Same as D, but for individual recording sessions and longer time windows.* ***(G)*** *Population decoding of drifting grating orientation (blue) and Shepard tone frequency (red) over time (same as A but with large temporal bins). Decoding aggregated across sessions was not significant overall (Wilcoxon signed-rank test, $p>0.05$ for virtually all time points).* ***(H)*** *Same as B, but for decoding of drifting grating orientation (blue) and Shepard tone frequency (red). Decoding performance was significantly above chance ($p<0.05$) in 5 out of 13 sessions for visual orientation frequency, and 5 out of 11 for tone frequency. This shows that in a subset of recordings, not only the presence of an auditory of visual change was encoded, but also the identity of the post-change stimulus.* ***(I)*** *Same as A, but for decoding rewarded versus unrewarded trials when aligned to first lick (note that licking activity/reward expectancy allows above-chance decoding before reaction time). Vertical dashed green line indicates the time point at which decoding accuracy reached half of the maximum accuracy (t=-0.25s).* ***(J)*** *Same as I, but for individual recording sessions and longer time windows.*

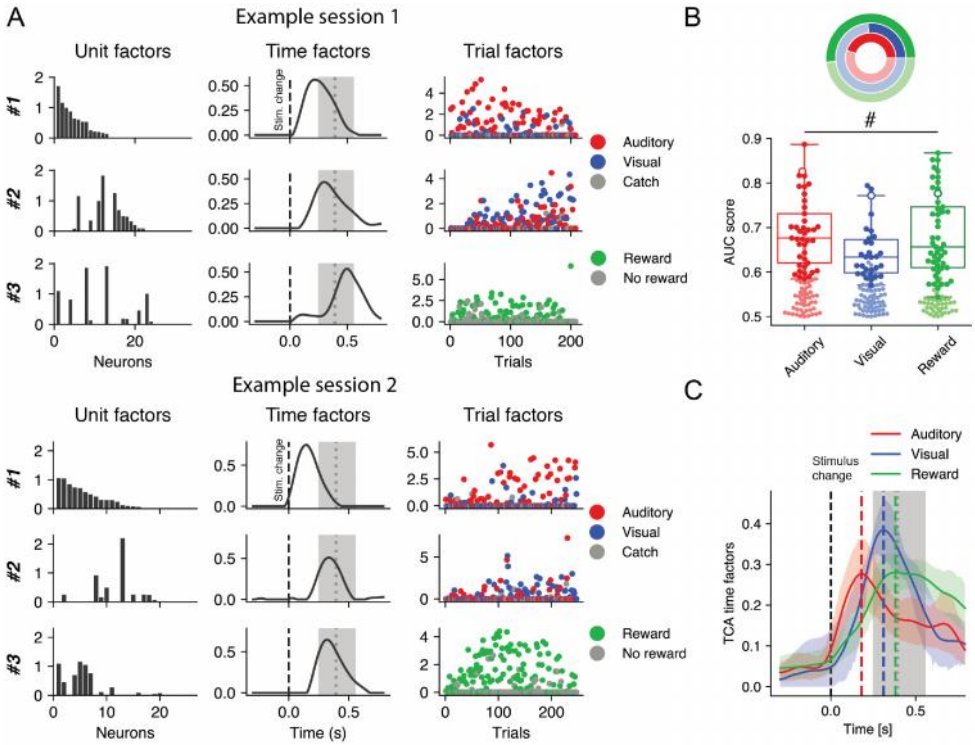
In addition to supervised decoding, we also used an *unsupervised* dimensionality reduction method (Williams et al., 2018) to test whether the dominant low-dimensional neural dynamics of PPC activity were task-relevant. This analysis showed that the distributed patterns of task-related neuronal activity were well described by latent components that closely corresponded to auditory and visual responses and decision-related behavior (Figure 5.5). Thus, well before reaction time, parietal population activity shows rich visual, auditory, and decision representations – including task-relevant information about which sensory modality shows a change – which are, theoretically, sufficient for downstream areas to execute the motor decision.

Sensory and task-related representations in PPC are not required for task performance

Altogether, when animals are trained to detect auditory and visual changes, PPC multiplexes sensory and task-related variables, thus suggesting a potential role in linking task-relevant auditory and visual inputs to adaptive decisions. To investigate the causal nature of this link, we optogenetically inactivated PPC bilaterally (by locally enhancing inhibition by parvalbumin-expressing (PV) interneurons). We verified that viral expression spanned PPC across the anteroposterior and mediolateral axes and across cortical layers (Fig. 5.6A-C) and that photostimulation excited a fraction of PPC neurons (putative ChR2-expressing PV-cells) while effectively suppressing ongoing activity (Fig. 5.6D-F). Optogenetic inactivation was less powerful in the deepest layers compared to more superficial ones (Fig. 5.6F), also in comparison to inactivation of V1 achieved using the same methodological approach (Fig. 5.6G). As opposed to the major silencing effect on PPC

itself, optogenetic inhibition of PPC had a slight excitatory effect on activity in V1 (Fig. 5.6H).

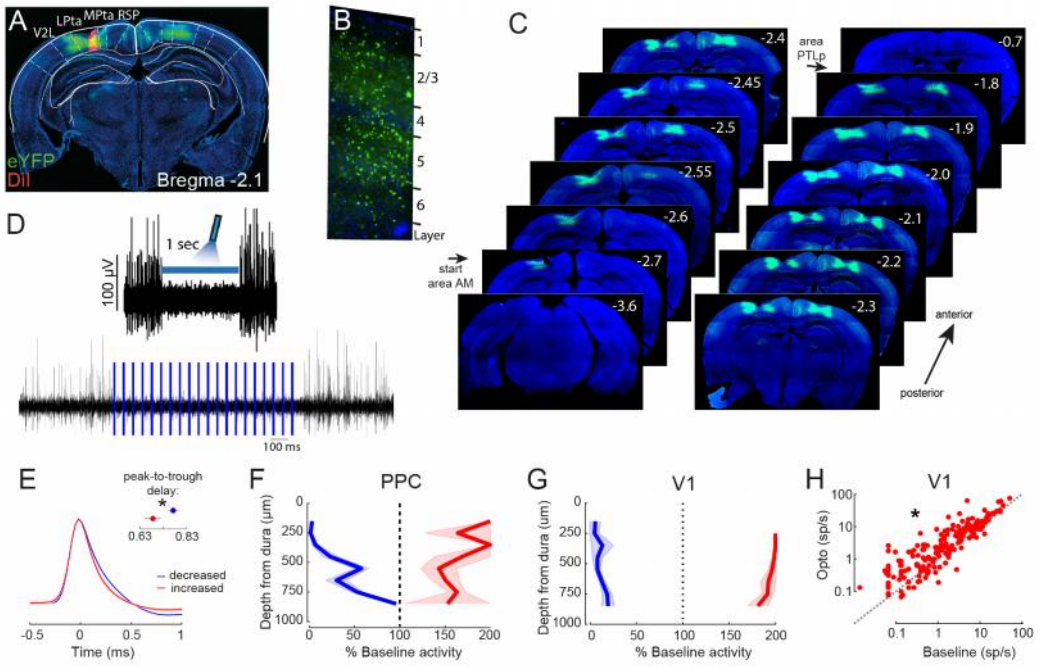
Figure 5.5: Latent dimensions of parietal population activity correspond to task-relevant components. *To capture, visualize, and quantify the dominant low-dimensional neural dynamics of population activity in PPC, we used an unsupervised dimensionality reduction technique, Tensor*



Component Analysis (TCA) (Williams et al., 2018). TCA decomposes population activity into a limited set of components, each of which corresponds to an assembly of cells with rapid, common within-trial dynamics expressed across a variable set of trials. ***(A)*** *Three selected components from a TCA analysis of two example sessions (8-component model; see Methods). Each row corresponds to a component (#1-3) that captures a subpopulation of neurons (unit factors, left column), that share a temporal response profile within the trial (temporal factors, central column), which is in turn expressed differently across trials (trial factors, right column). All factors are unitless. Neurons are ranked based on their contribution to the first plotted component, then the second, etc. The dotted line and the shaded grey (middle panel) indicate the median and interquartile range of reaction times, respectively. The TCA decomposition is performed with no information about trial types, which are colored a posteriori. Nonetheless, the identified latent components show selective expression in auditory (#1), visual (#2), and hit trials (#3) in both example sessions, showing that the main dimensions of population activity are task-related.* ***(B)*** *AUC scores measure how individual components are selectively expressed in auditory versus catch trials (red), visual versus catch trials (blue), and rewarded versus not rewarded trials (green). Components whose AUC scores were significantly larger than chance (tested against random trial permutations, $p<0.05$) are marked by darker dots. Boxes indicate the quartile and 1.5 times the interquartile range of the significant AUC scores. White-filled dots locate the AUC scores of the three components shown in the top TCA decomposition of (A). Donut plot shows the fraction of components that have a significant AUC score for each contrast, across all recordings. '#' BF<1/3.* ***(C)*** *Averaging the time*

factors of all selective components from different sessions (i.e. those with a significant AUC score, corresponding to the components marked in darker dots in B) showed a distinct temporal profile with a progression of first auditory, then visual, and lastly decision components. Shaded bands indicate 95% confidence intervals. Vertical dashed lines mark the peak time of the averaged temporal factors.

Figure 5.6: Optogenetic inactivation of Posterior Parietal Cortex. (A) *Coronal section showing viral expression in bilateral PPC and the electrode tract stained with DiI for simultaneous physiological validation.* ***(B)*** *Viral infection was present throughout layers of PPC. No evidence was obtained for infection of deep pyramidal neurons with large somata, as reported for S1 (Tanahira et al., 2009).* ***(C)*** *Series of coronal sections from one mouse showing the extent of viral infection of PPC along the anteroposterior and mediolateral axis, cf. (Paxinos and Franklin, 2004). Numbers denote approximate anteroposterior offset from Bregma. When aligning to the Allen Mouse Brain Atlas, the first expression starts in the Anteromedial visual cortex (AM) and ends in the Posterior Parietal*



Association area (PTLp). ***(D)*** *We inactivated PPC by local activation of Parvalbumin-expressing interneurons and photostimulation with an optic fiber (2-10 mW, 10 ms pulses at 25 Hz) directly over the thinned skull. Top trace shows the high-pass filtered voltage (>500 Hz) at an example electrode site (recorded at approximately 450 µm below dura) with effective inhibition of spiking activity (i.e., multi-unit activity, MUA) upon photostimulation. Bottom trace shows a close up of individual pulses during 1 second photostimulation.* ***(E)*** *The average waveform was narrower for PPC neurons that showed increased (red) versus decreased (blue) firing rate upon photoinactivation of PPC. This is in line with putative narrow-spiking PV cells being recruited by photoinhibition to suppress broad-spiking pyramidal cells. Inset shows a significant difference in the peak-to-trough delay (0.81 vs 0.70 ms, 90 vs 38 neurons, BF=4.47). *BF>3.* ***(F)*** *Effectiveness of optical manipulation across cortical depth is shown as the percentage of the baseline firing rate. Neurons were separated into inhibited (n=95 neurons; blue) and excited (n=39 neurons; red) populations. We clipped the range of modulation ratios at 200% (some neurons were strongly recruited by photostimulation). Lines and shading show mean ± SEM.* ***(G)*** *Same as F, but for V1 inhibition (n=150 inhibited neurons in blue, 68 excited neurons in red). Compared to PPC,*

optogenetic inhibition in V1 was more effective in the deeper layers. ***(H)*** *To investigate the spatial extent of PPC inactivation we analyzed V1 activity during PPC photostimulation. Optogenetic inhibition of PPC had a slight excitatory effect on activity in V1 (Bayesian t-test photostimulation vs baseline, n=213 V1 neurons, Cohen's d: 0.41, BF = 4.08), as opposed to the major silencing effect on PPC itself. At the minimum, this suggests PPC inactivation does not affect V1 in a major way, in line with the extent of viral expression seen in C. *BF>3.*

Surprisingly, optogenetic silencing of PPC had significant effects on neither visual nor auditory change detection, with evidence for the absence of an effect (Keysers et al., 2020) in both visual and auditory conditions (Figure 5.7A-C). We additionally inactivated V1 in a separate cohort of animals. Bilateral inactivation of V1 strongly reduced detection of visual orientation changes, but not auditory frequency changes, consistent with the primary role of V1 in visual feature processing (Glickfeld et al., 2013b; Resulaj et al., 2018; Zatka-Haas et al., 2021).

PPC inactivation could affect other aspects of behavioral performance, despite overall preserved visual and auditory hit rates. However, we found that all tested aspects of behavioral performance persisted with PPC inactivation. First, we found no effect on reaction times (Figure 5.7D). Second, we tested the hypothesis that PPC inactivation could be restricted to a subset of animals showing a particular behavioral strategy. However, we found both small increases and decreases in task performance in individual mice and no relationship between changes in task performance and the bias to report visual or auditory stimuli (Figure 5.7E,F). Third, even though detection of stimuli at perceptual threshold – those most sensitive to perturbation – was preserved (Fig. 5.7A-C), psychometric performance could potentially be affected in other subtle ways. However, when we inactivated PPC at various levels of auditory and visual stimulus difficulty levels, hit rates were unaffected (Figure 5.8A). We fit each session with a psychometric version of our signal detection model and found no evidence for an effect - but also no evidence for the absence of an effect - on sensitivity, threshold, and bias (Figure 5.8B, C).

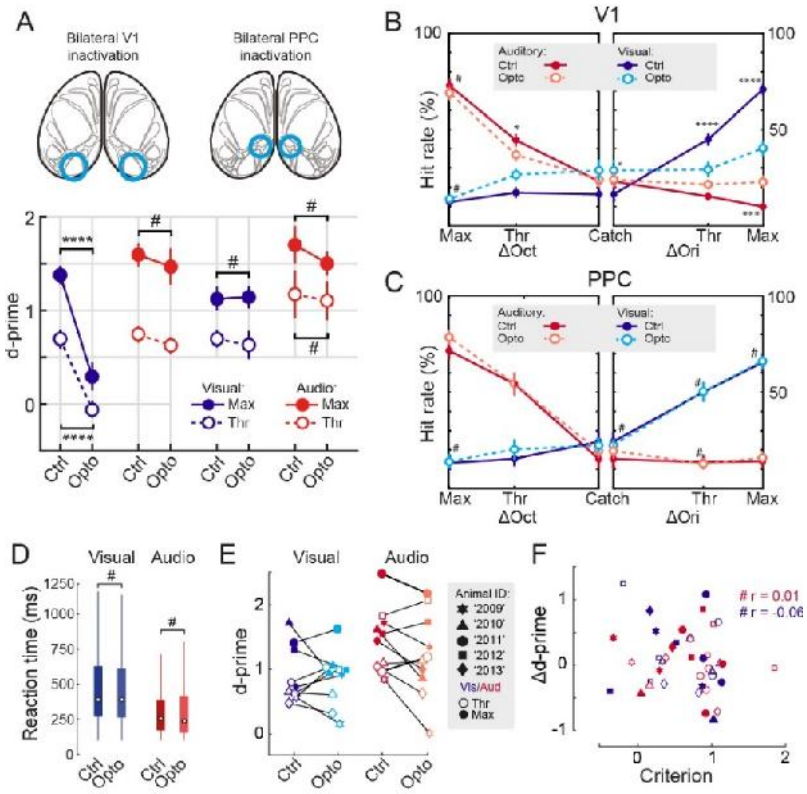


Figure 5.7: Optogenetic inactivation of Posterior Parietal Cortex does not affect visual or auditory change detection. (A) *Bilateral V1 inhibition reduced visual detection performance as quantified by d' with a two-alternative signal detection model (Bayesian paired t-test, 19 sessions; Threshold visual changes: Cohen's d = 1.162, BF=135.55; Maximal visual changes: Cohen's d = 1.251, BF=593.51), with no effect on auditory performance (Threshold auditory changes: Cohen's d = 0.231, BF=0.49; Maximal auditory changes: Cohen's d = 0.132, BF=0.30). Bilateral PPC inhibition had no effect on either visual or auditory performance (Bayesian paired t-test, 15 sessions; Threshold visual changes: Cohen's d = -0.197, BF=0.34; Maximal visual changes: Cohen's d = -0.110, BF=0.30; Threshold auditory changes; Cohen's d = -0.178, BF=0.27; Maximal auditory changes: Cohen's d = 0.417, BF=0.32).* ***(B)*** *Behavioral response rates averaged over V1 inactivation sessions for control and photostimulation trials. Error bars show SEM across sessions.* ***(C)*** *Same as B, but for PPC silencing, where behavioral hit rates are unaffected.* ***(D)*** *PPC inactivation did not affect reaction times. Boxplot show median and interquartile ranges. Auditory hits: 257 ms (control, median reaction time), 238 ms (opto), 1454 vs 915 trials, Cohen's d = -0.036, BF=0.033; Visual hits: 391 ms (control), 392 ms (opto), 1115 vs 735 trials, Cohen's d = -0.001, BF=0.016. Outliers were omitted from the plot for visual clarity.* ***(E)*** *Average d-prime with and without PPC inactivation for each animal. Same data as the right part of A.* ***(F)*** *The change in d-prime by PPC inactivation was not related to the threshold for visual or auditory report, i.e. the criterion parameter in signal detection theory (Bayesian correlation; visual: r=-0.06, BF=0.18; auditory: r=0.01, BF=0.17). Each dot is a session. *, **, ***, ****, indicate BF > 3, 10, 30, 100, respectively; # is BF<1/3 and denotes evidence for the absence of an effect.*

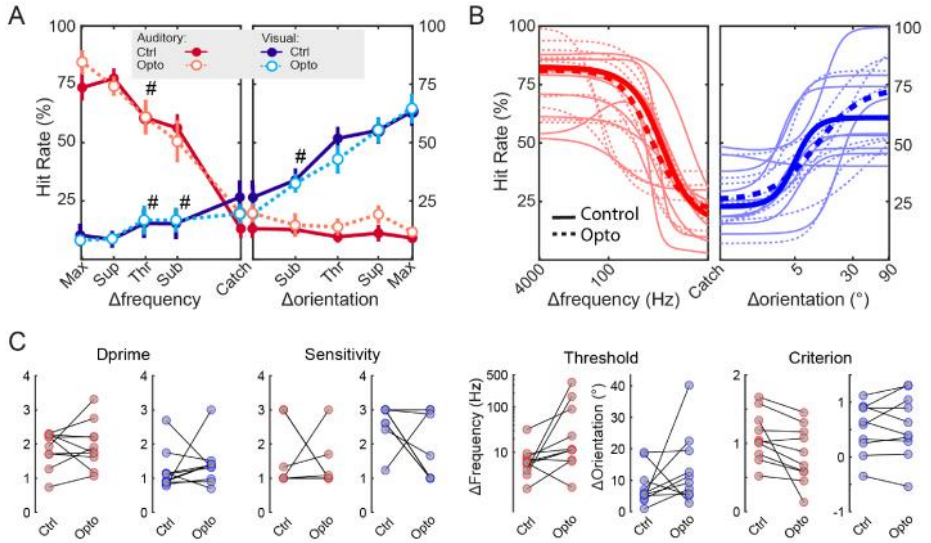


Figure 5.8: Inactivating PPC does not affect psychometric parameters in the audiovisual detection task. ***(A)*** *Same as Figure 5.7C, but now for PPC inactivation during sessions with four levels of auditory and visual change (n=9 sessions).* ***(B)*** *We fitted a psychometric two-alternative signal detection model on control and photostimulation trials of individual sessions separately. Thin lines show single session fits, thick lines show the median fit. Based on the same data as in A.* ***(C)*** *For each of the eight model parameters we found no evidence for an effect on fitted parameter value comparing control and photostimulation trials with a Bayesian paired t-test (n=9 sessions): Auditory parameters: d', Cohen's d = -0.196, BF=0.363; sensitivity: Cohen's d = 0.668, BF=1.469; threshold: Cohen's d = -0.534, BF=0.898; criterion: Cohen's d = 0.761, BF=2.108; Visual parameters: d': Cohen's d = -0.160, BF=0.344; sensitivity: Cohen's d = 0.134, BF=0.333; threshold: Cohen's d = -0.419, BF=0.617; criterion: Cohen's d = -0.154, BF=0.342. For 4 of these parameters BF was close to (but not below) 1/3 suggesting evidence of absence of an effect.*

Behavioral choice is history-dependent, but this effect is not dependent on posterior parietal cortex

Previously, PPC was shown to mediate history-dependent effects on behavioral choice (Akrami et al., 2018; Hwang et al., 2017). To investigate this, we constructed a multinomial logistic regression model that uses information about the current and previous trial to predict the animal's choice (3 response options: lick to visual spout, lick to auditory spout or no lick). To test the contribution of individual sensory and behavioral features, we constructed ten models with different regressor combinations (Fig. 5.9A). Data was fit on concatenated sessions of individual animals and model predictions were evaluated on held-out test data (3-fold cross-validation). Relative to a null model, including sensory history improved model performance slightly, while including reward history and choice history had larger effects (Fig. 5.9B). Including information about the sensory stimulus of the current trial (amount of visual and auditory change) boosted performance significantly, confirming animals mostly base their choice on the current stimulus.

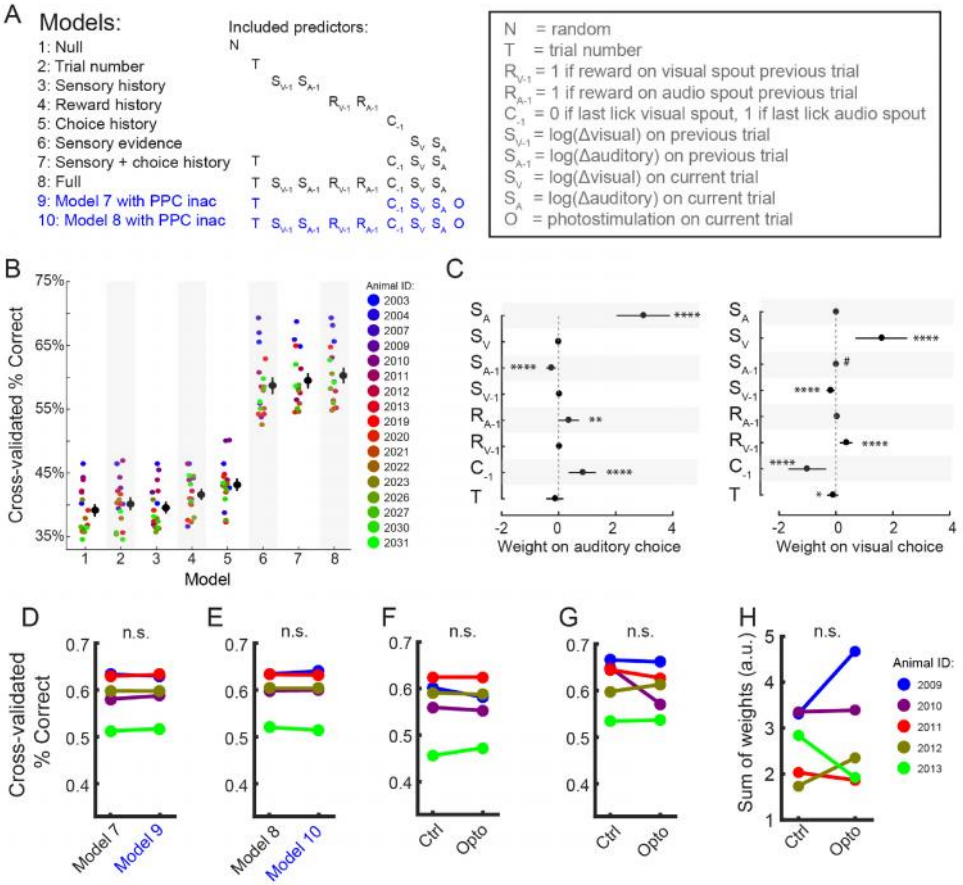


Figure 5.9: Behavioral choice is history-dependent, but this effect is not dependent on posterior parietal cortex. ***(A)*** *Overview of ten models with different combinations of regressors. A null model served as a baseline comparison, and models 2-6 included different single predictors, while models 7-10 contained multiple predictors.* ***(B)*** *Cross-validated performance (% correctly predicted choices on held-out test data) for the 8 model variants (without photoinactivation of PPC). Each dot is the average performance over 3 folds for one animal. Models differed in their predictive performance (Bayesian Anova, Bayes factor relative to intercept only=5.6x10^{32}). Models 2-6 all performed better than the null model (Trial number: Cohen's d: 0.67, BF=41.92; Sensory history: Cohens' d: 1.34, BF=4.10; Reward history: Cohen's d: 1.27, BF=577.24; Choice history: Cohen's d: 3.51, BF=350.15; Sensory evidence: Cohen's d: 4.25, BF=6,6x10^{7}).* ***(C)*** *Weights of the different predictors on probability ratios of licking to auditory and visual spouts relative to no-lick in the full model (Model 8). Errorbars indicate mean ± SEM. *,**,***,*** BF>3, 10, 30, 100 respectively, # BF<1/3.* ***(D)*** *The effect of PPC inactivation on choice behavior was tested in four different ways. First, we compared model performance with and without including photostimulation as a regressor in the sensory and motor history model (Model 7 vs 9, Bayesian t-test; Cohen's d = -0.519, BF=0.643). Each dot is one animal.* ***(E)*** *Same as D, but for the full model with and without photostimulation (Model 8 vs 10, Bayesian t-test; Cohen's d = -0.006, BF=0.397).* ***(F)*** *Second, as in (Hwang et al., 2017) we compared model performance when training on control trials and testing on control or photostimulation trials (Bayesian t-test; cross-validated correct trials on control versus opto): Cohen's d = 0.217, BF=0.437).* ***(G)*** *Third, as in (Akrami et al., 2018) we tested model performance (Model 8) when training and testing separately on control or photostimulation trials only (Bayesian t-test; cross-*

validated correct trials, control versus opto: Cohen's d = 0.454, BF=0.582). ***(H)*** *Finally, as in (Akrami et al., 2018) we compared the sum of motor and reward weights when training and testing separately on control or photostimulation trials only (Bayesian t-test; sum of history weights control vs opto: Cohen's d = -0.214, BF=0.436).*

We investigated the regressor weights in the full model, with positive auditory or visual weights indicating a higher probability of licking the auditory or visual lick spout, respectively, relative to not licking (Fig. 5.9C). Trial number had a small negative weight on both lick probabilities, in line with growing satiety. The other weights were modality-specific. Auditory and visual stimuli on the current trial had strong positive weights, sensory history a slight negative weight, and previous reward a positive weight. The previous choice had opposite weights suggesting animals repeated their choice.

We next tested the effect of PPC inactivation on history-dependent behavioral choice in four different ways, but found no evidence for effects of PPC inactivation on how sensory information of the current trial is used, nor how trial history affects behavior (Fig. 5.9D-H) (Akrami et al., 2018; Hwang et al., 2017). In sum, although we found that behavioral choice was influenced by reward and choice history (but not sensory history), these history effects were unchanged by PPC inactivation.

Lastly, we presented a subset of multimodal trials in which both modalities changed simultaneously, which implies a conflicting situation (i.e., mice could lick either left or right, based on which sensory modality prevails). In contrast to Song *et al.* (2017), who found that PPC inactivation shifted behavioral choice from auditory to visual report, we observed no such effect (Figure 5.10). Therefore, despite robust encoding of sensory- and task-related variables well before response onset, we found no causal relevance of PPC in audiovisual change detection.

Discussion

We tested the hypothesis that PPC is required to arbitrate between sensory modalities in a change detection task that required mice to segregate and identify sensory modalities. Training mice on this task led to qualitatively rich and heterogeneous auditory, visual, and decision-related activity changes in PPC (Figure 5.1-5.5), which was however not causal to performing the task (Figure 5.7-5.10).

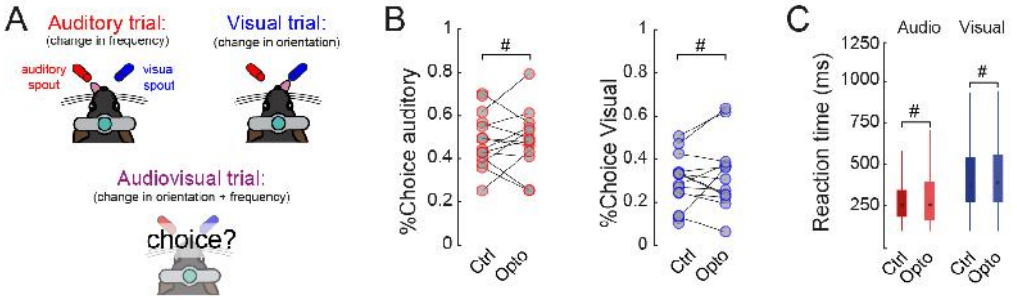


Figure 5.10: Behavioral choice during audiovisual trials is preserved during PPC inactivation. (A) *As visual and auditory feature changes were associated with rewards at different lick spouts and thus different motor actions, the modalities acted as competing inputs. A simultaneous change presents the animal with a conflicting situation (i.e., mice could lick either left or right, based on which sensory modality prevailed). A reward was given for licking either spout.* ***(B)*** *We tested whether PPC inactivation specifically affected the auditory or visual choice fraction in conflict trials; this was not the case (Bayesian t-test, N=15 sessions; Auditory choice: Cohen's d=0.21, BF=0.289; Visual choice: Cohen's d = -0.404, BF=0.30). '#' BF<1/3 and denotes evidence for the absence of an effect.* ***(C)*** *We found no effect of PPC inactivation on reaction times only during conflict trials. Auditory choice trials: 253 ms (control), 255 ms (opto), 507 vs 290 trials, Cohen's d = -0.027, BF=0.018. Visual choice trials: 368 ms (control), 390 ms (opto), 284 vs 211 trials, Cohen's d = -0.031, BF=0.018. Boxplots indicate median and interquartile ranges.*

Previous studies reported a similar absence of behavioral consequences despite strong decision-related activity in PPC (Erlich et al., 2015; Hanks et al., 2015; Katz et al., 2016; Licata et al., 2017; Raposo et al., 2014). Our results further reveal a surprising dissociation between rich representations in PPC (preceding reaction time) and causal involvement in task performance. This is especially surprising given the dense interconnection of PPC with both primary sensory cortices (V1 and A1) and nearby associative areas such as anterolateral visual cortex (AL) (Meijer et al., 2020) and premotor areas such as supplementary motor cortex, which show multisensory task correlates and are causally involved in task performance (Allen et al., 2017b; Barthas and Kwan, 2016; Coen et al., 2021; Erlich et al., 2015). Therefore, in spite of its uniquely suited anatomical and functional properties, at the interface between visual and auditory processing and premotor cortices, PPC is required for neither auditory nor visual change detection. A different result might have been obtained if visual and auditory had to be integrated. Another possible explanation, however, is that sensory and decision-related activity in PPC may reflect secondary processes not directly subserving behavioral decision-making in our task, or functional processing of an order of complexity that is not required to solve the task, as this may be solved by lower-order (e.g. subcortical) structures. Interestingly, the large increase in auditory responses that we reported in trained mice conforms with similar reports in rodent association cortices (Meijer et al., 2020; Raposo et al., 2014; Zhong et al., 2019), but is different from what has been reported in primates (Cohen, 2009; Grunewald et al., 1999). In particular, limited auditory responses have been reported in primate area LIP upon engagement in an auditory task (Grunewald et al., 1999). This may indicate different circuitry for auditory processing in rodent vs. primate PPC, but also the fact that primate experiments were performed in an area (LIP) that bears no homologue in rodents. The increase in visual and auditory responses that we observed could result both from extensive

behavioral training as well as active task engagement, as both were confounded in our experimental comparison between trained and naïve mice.

Considering the broader literature on rodent PPC, under which task conditions would parietal circuitry be causally involved? In recent studies, stimuli and context were fixed while task demands were varied, and PPC - amongst other regions - became necessary only when evidence accumulation, working memory, or in any case an extension in the time scale of the task were required (Arlt et al., 2021; Harvey et al., 2012; Pinto et al., 2019). Our results fit with this view and suggest that PPC is dispensable for tasks that have fixed one-to-one sensorimotor mappings (but see Arlt et al. 2021). Along these lines, PPC is thought to participate in a larger cortical network constructing contextual representations of multimodal inputs. These representations likely serve to maintain information across a delay (working memory) or across trials (history effects) or integrate information with previous experience (learning) and are not crucial for the detection of sensory stimuli. An interesting parallel interpretation of the richness of neural PPC correlates in the absence of a causal role in audiovisual change detection relates to the involvement of the human PPC in hemineglect (Driver and Vuilleumier, 2001; Kerkhoff, 2001; Vallar, 1998), pointing to a function of the PPC in generating the multisensory spatial survey that we associate with consciousness and spatial attention (Pennartz, 2015; Pennartz et al., 2019). Intriguingly, our results shows that PPC displays classical neural correlates of consciousness (Koch et al., 2016), without being causally involved in the reportability of sensory stimuli. This indicates the need to critically evaluate the potential relevance of neuron-level correlates of sensory detection in the context of consciousness research – but see (Nieder et al., 2020; Vugt et al., 2018) for recent studies in which such correlates are interpreted as markers of consciousness. Overall, based on this functional angle, PPC activity may become behaviorally relevant when utilization of this contextual information is necessary, for example in guiding decisions based on unfamiliar sensory stimuli or during learning (Wander et al., 2013; Wilber et al., 2017; Zhong et al., 2019), as also reported in primary somatosensory cortex (Hong et al., 2018) or in a more complex task requiring a multisensory spatial survey of the subject in its environment.

Another possible interpretation is that choice-related activity in PPC may reflect - at least to some extent - movement variables (Mimica et al., 2018; Whitlock et al., 2012), which are increasingly reported as affecting widespread cortical circuits in rodents (Musall et al., 2019; Salkoff et al., 2020; Stringer et al., 2019). Indeed, movement variables did explain a fraction of the variance across the PPC population activity (Figure 5.3D). Alternatively, the modality identification task tested here may be solved by a distributed network of nodes in which the PPC participates in a causally redundant manner or that PPC's role is redistributed through fast-acting compensatory mechanisms (Mejías and Wang, 2022; Sigler et al., 2009). The inconsistency between our reported absence of an effect of PPC inactivation on auditory dominance with Song et al. (2017) may lie in the fact that we used instantaneous sensory changes, which do not result in an extended temporal window of conflicting signals. Finally, a potential concern is that our protocol for optogenetic inactivation spared a limited functionality of the PPC in its deep layers sufficient for full task performance (Figure 5.6F). Although viral infection expression encompassed PPC across the full mediolateral axis of PPC and reached the deepest layers (Figure 5.6A-C), and we effectively suppressed visual perception using the same experimental approach in V1, we cannot exclude that a portion of PPC might have been spared and might have been sufficient to prevent behavioral impairment.

In conclusion, we showed that neuronal responses in mouse PPC during audiovisual change detection, in spite of being multisensory and anticipating task-related responses, are not causally related to task performance. This sheds light on the function and architecture of parietal associative circuits and emphasizes the importance of cautiously interpreting the causal relevance of neural activity for task performance.

Acknowledgments

We thank D. Sridharan for providing code for the multi-alternative detection model; A. Williams for code for Tensor Component Analysis; C. Rossant, collaborators, and members of the Cortex Lab (UCL) for Klusta and Phy spike sorting software; A. Mathis and collaborators for DeepLabCut software; This work was supported by the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement 720270 (Human Brain Project SGA1) to C.M.A.P., Grant Agreement 785907 (Human Brain Project SGA2) and 945539 (Human Brain Project SGA3) to C.M.A.P. and U.O., by the FLAG-ERA JTC 2015 project CANON (co-financed by the Netherlands Organization for Scientific Research – NWO) to U.O., and by the FLAG-ERA JTC 2019 project DOMINO (co-financed by NWO) to U.O.

Author contributions

	MOL	PM	CMAP	UO
Conceptualization	X		X	X
Formal Analysis	X	X		
Investigation	X			
Methodology	X			
Resources	X			
Software	X			
Writing – Original draft	X			
Writing – Reviewing and editing	X	X	X	X
Supervision			X	X
Funding & Acquisition			X	X

Table 1: Absolute and shuffled population decoding accuracy

Decoding variable:	Trial contrast:	Shuffled accuracy:
Auditory presence	Auditory trials vs. catch trials	0.603 (0.521 - 0.660)
Visual presence	Visual trials vs. catch trials	0.588 (0.519 - 0.633)
Modality	Auditory vs. visual trials	0.503 (0.503 - 0.507)
Reward (stim-aligned)	Rewarded vs. unrewarded trials	0.548 (0.525 - 0.579)
Auditory frequency	Post-change frequency AB vs. CD	0.528 (0.497 - 0.584)
Visual orientation	Post-change orientation AB vs. CD	0.527 (0.509 - 0.578)

This table reports the absolute decoding accuracy, in contrast to the performance above accuracy on shuffled trial labels as presented in the main text and Figure 5.4. For each of the decoded variables (first column), it lists which two sets of trials were contrasted (second column), the median and interquartile range of the cross-validated classification accuracy on shuffled trial labels (third column), and accuracy when trained and tested on the true trial labels (fourth column) and to which panel from Figure 5.4 the analysis corresponds to (fifth and last column). Decoding accuracy is performed at t=0.25s (time bin extending from 0 to 0.5s).

Chapter 6: General Discussion

Summary of results

This thesis investigated how visual processing is impacted by different contextual factors across four chapters. Context was defined in a broad sense as modulation by sources external to visual cortex that convey information other than direct visual input of retinal origin.

In Chapter 2, we studied how higher visual areas (HVAs) shape activity in primary visual cortex (V1). We found that optogenetic inactivation of two HVAs with different functional properties similarly affected basic visual response properties of V1 in both awake and anesthetized conditions. These areas enhanced stimulus responsiveness to moving bars by more strongly amplifying weaker compared to stronger sensory-evoked responses. This degraded feature selectivity, but this amplification of responses to stimuli that would otherwise not induce large responses enables more robust responses to weak visual stimuli. Feedback from higher-order visual areas thus selectively amplifies weak sensory-evoked activity in V1 and determines basic properties of visual processing.

In Chapter 3, we investigated whether the function of V1 during perceptual decision making is dependent on task demands in other modalities. Specifically, we tested whether V1 functions as a fixed visual processing step or as a node in a network in which the contribution of V1 might change depending on the task needs. We found pronounced V1 activity changes in a late stage of the V1 response in task settings where vision was task-relevant. The onset of this late response and reaction times were delayed in task settings with additional auditory or tactile relevance. Late-onset optogenetic inactivation of V1 only affected visual detection in multisensory contexts. In other words, although the demand to report identical visual stimuli was unchanged, the addition of an extra modality prolonged reaction times and extended the causal involvement of V1. We interpreted these results as an indication that the temporal window transforming the same sensory inputs into a perceptual representation and motor decision requires substantial recurrent interplay between cortical areas, which depends on multisensory task context.

In Chapter 4, we investigated the nature of sound-evoked activity in visual cortex. Sound-evoked activity in V1 is thought to reflect auditory sensory inputs to the visual system, mediated in part through early (subcortical or cortical) auditory pathways to V1. However, sounds can evoke body movements that are known to also modulate the visual cortex, suggesting that previously reported sound-evoked activity in V1 rather reflects secondary behavioral modulations. To address this controversy, we disentangled auditory-related from motor-related activity using multi-area recordings, task manipulations, pharmacological interventions, and optogenetics. Sound-evoked activity consisted of both auditory and motor components, with distinct spatiotemporal profiles. Short-latency auditory-related activity originated from auditory cortex and was found in superficial and deep layers of V1. However, sounds also evoked rapid orofacial movements accompanied by distinct motor-related activity distinguishable as early as 60-100 ms. These separate components co-existed in visual cortex in a way that, during conflicting audiovisual inputs, the ability of visual cortex to represent visual features was preserved and visual orientation coding stayed intact.

In Chapter 5, we tested the hypothesis that Posterior Parietal Cortex (PPC) is required to arbitrate between the noisy inputs from the two different modalities and help decide in which modality a sensory change occurred in our newly developed audiovisual change detection task. Despite extensive single-neuron and population-level encoding of task-

relevant visual and auditory stimuli, as well as upcoming behavioral responses, optogenetic inactivation of PPC did not affect task performance.

Contextual modulation as predictive coding

We found that V1 activity was affected by feedback from higher visual areas (Chapter 2), auditory signals (Chapter 4), and orofacial movements (Chapter 3-4). How can this diversity of non-visual signals in V1 be interpreted in a unified way of how vision works? One candidate framework is that of predictive coding. This broad theory states that what we directly perceive are not the direct sensory inputs, but rather the inferred cause of the sensory data; in other words, we perceive that what we infer to have most likely caused the specific pattern of inputs across the senses (Gregory, 1980; Helmholtz, 1867; Lee and Mumford, 2003). How these inferred causes relate to external signals constitutes an internal model. This model serves to efficiently interpret new sensory inputs, as well as continuously generate predictions of the world and body in space or time. One can both infer and learn the internal representations using deviations from what was predicted, what is referred to as prediction errors.

Amongst the many proposed advantages of such a coding scheme is its efficiency. For instance, the visual domain is full of strong spatial and temporal correlations. Patterns in natural scenes are often predictable across space (e.g. the horizon from left to right) and time (remaining roughly the same from one moment to the next). An occluded part of a tree can be predicted based on the rest of the tree, and the future trajectory of a bird from its past flight. Much of visual input can therefore be explained by nearby visual inputs (spatial predictions) or past inputs (temporal predictions). A processing strategy based on removing these redundancies has been argued to be highly efficient, given that one needs to only process the errors rather than the full image. A further extension involves not only using current spatial and temporal correlations, but also predictions based on prior experience (Attneave, 1954; Barlow, 1961; Knill and Richards, 1996; Lewicki, 2002; Mumford, 1992; Olshausen and Field, 1997; Srinivasan et al., 1982). Under this view, the main function of the visual cortex is to learn and maintain an internal model of the visual world, and visually evoked activity mainly corresponds to the processing of unexpected changes in the visual environment to guide updating of the internal model. In addition to processing the errors, predictive representations also have to be maintained that track current states (Keller and Mrsic-Flogel, 2018; Pennartz et al., 2019; Spratling, 2017). How is the myriad of context-dependent inputs to V1 to be interpreted within this framework?

Higher-order feedback - First, feedback from higher to lower areas plays an essential role in most hierarchical network implementations of predictive coding. How feedback projections could contribute to encoding images has been explored in a seminal modeling framework (Rao and Ballard, 1999). In their 2-layer neural network sensory inputs from an input layer were used by a higher layer to generate a predictive representation of the actual sensory input. The predictive representation is fed back to the input layer via recurrent projections and updating this network explained extraclassical receptive field properties like end-stopping and contextual modulations (Rao and Ballard, 1999). In the mouse visual cortex, feedback connections have been shown to provide contextual modulation to V1 neurons of relevant features outside their receptive field (Keller et al., 2020a; Kirchberger et al., 2021). Especially L2/3 neurons show diminished responses if their input is highly predictable from the surround and the suppression of predictable responses is mediated by the VIP disinhibitory circuit (Keller et al., 2020a). In the experiments presented in Chapter 2, we used a single bar moving across the visual field.

From the past trajectory, future visual input can be predicted for V1 neurons that represent spatial locations just ahead of the bar. HVAs, with their larger receptive fields and longer timescales of activity (Runyan et al., 2017; Siegle et al., 2021), are poised to convey these predictions to V1. Our finding that HVA inactivation selectivity affected V1 responses as the predictable moving bar is about to enter the receptive field (Fig. 2.3) is in line with a specific role for HVAs in modulating V1 based on input predictability. A circuit implementation would involve HVA projections to V1 neurons (possibly through interneurons) that represent retinotopic locations spatially ahead of the direction of the bar. This interpretation is inconsistent however with an earlier study that found that HVA feedback preferentially targets V1 neurons that represent the part of visual space that is actually slightly behind what the higher-order neurons have information about (Marques et al., 2018). This inconsistency might result from our silencing not just HVA to V1 feedback but rather the whole area, and leaves open other circuit implementations of predictive processing. Namely, whether and how hierarchical networks use predictive computations critically depends on the supposed circuit implementation of the algorithm (Keller and Mrsic-Flogel, 2018; Pennartz et al., 2019; Teufel and Fletcher, 2020). For example, a different conceptual interpretation would arise in a predictive circuit model where representations are propagated forward and the errors are propagated backward (Heeger, 2017).

Audition - Auditory signals in the visual cortex have been explained within the PC framework as well. Just as visual scenes are full of spatial and temporal correlations, so are correlations between auditory and visual signals abundant. Learning the structure of these relationships allows one to predict signals from one domain to the other. The sight of a dog opening his jaws is associated with hearing barking sounds, and the sound of an overhead helicopter predicts seeing one. Here, 'prediction' can be understood in a temporal sense (stimulus X predicts what will happen next) or may as well pertain to 'predicting the present' (stimulus X at time t influences the perception of another stimulus Y at or around time t). One main function of auditory-to-visual inputs might then be to facilitate visual inference based on past audiovisual experience. Indeed, audio-visual experience can cause neuron-specific reconfigurations in V1 (Knöpfel et al., 2019). Furthermore, after repeated sequential presentation of the same auditory and visual stimulus pair, the responses of V1 neurons to the same visual stimulus can be reduced by the auditory stimulus (the prediction error is smaller because the visual stimulus is expected based on the preceding sound) (Garner and Keller, 2022). Auditory stimuli have further been reported to improve visual feature representation (Ibrahim et al., 2016; Williams et al., 2021). These auditory modulations are therefore interpreted to facilitate cue inference in classic multisensory integration frameworks (which can be reformulated as a case of predictive inference without strong priors). In Chapter 4, however, we reported low-latency auditory signals in V1 not only in naive animals, but also in animals trained to segregate rather than integrate audiovisual inputs. These signals (jointly with motor-related activity) did not boost or degrade visual orientation coding (Fig. 4.6), but evoked spiking in a separate set of V1 neurons that were not visually driven (Fig. 4.3e). Our results suggest that auditory inputs do not merely serve immediate inference of present visual features, moving beyond tuning curve modulations. Rather, they might act - also in contexts without immediate inference - as an early predictor (present earlier than visual stimuli, Fig. 4.3b), enabling V1 to generate a prediction about what we will see, or are currently seeing.

Motor movement – In the predictive coding framework, body movement signals in sensory areas are readily explained. Movement complicates even the most basic sensory inference that would be trivial for a stationary observer. For example, during vision, head and eye movements transform the image of a static visual scene into a complicated and

dynamic retinal input that depends on the precise gaze and eye movements. Determining features of a stationary input, such as the orientation, suddenly becomes nontrivial. Distinguishing self-generated stimulation ('reafference') from stimulation driven by the outside world ('exafference') is therefore fundamentally ambiguous and can only be resolved using internal signals about self-generated motor commands (termed *efference copy* or *corollary discharge*) (Crapse and Sommer, 2008; Cullen, 2019; Schneider and Mooney, 2018; Schneider et al., 2014; Sommer and Wurtz, 2008). Accurate visual inference about the external environment necessitates corrective computations using these predicted sensory consequences of motor commands.

We found that in task-trained animals (UST and MST cohorts; Ch. 3), correct reports but also false alarms were associated with a large increase in activity in a significant fraction of V1 neurons. Regression modeling and carefully observing body movements showed that activity during visual, but also auditory trials partly related to motor activity (Ch. 3-5). This increase in activity was also seen in PPC (Fig. 5.2 and 5.3), and to a lesser extent in auditory cortex (AC) (Fig. 4.3). Moreover, this brain-wide distributed activity was observed to a lesser extent in naive animals. These results align with a series of recent studies that show a distributed coding of goal-directed movements in trained animals and coding of ongoing movements and behavioral variables in naive animals (Allen et al., 2017b; Musall et al., 2019; Salkoff et al., 2020; Stringer et al., 2019). Surprisingly, in our data as well as in the cited studies the variability of V1 firing activity explained by these orofacial movements was comparable to or exceeded that explained by visual input (Fig. 4.3d). One interpretation of such pervasive motor signals is that perceptual inference throughout cortical sensory systems fundamentally relies on access to, correcting for and/or canceling self-generated movements (Kaplan and Zimmer, 2020; McCormick et al., 2020). Indeed, locomotion is accompanied by frontal-to-V1 signals conveying predicted visual flow and V1 L2/3 neurons incorporate these predictions and predominantly signal deviations from expected visual flow (Attinger et al., 2017; Fiser et al., 2016; Keller et al., 2012; Leinweber et al., 2017). Furthermore, V1 neurons are tuned to specific head and eye movements depending on ambient light, i.e. depending on whether visual changes are expected (Bouvier et al., 2020; Guitchounts et al., 2020). Motor-related activity in sensory areas can thus be interpreted as sensory-predictive signals.

Perceptual outcome - The hit-related activity that we observed was also partly explained by trial outcome (Fig. 3.2c). In our behavioral task, as well as many other tasks used in the field, hit trials are associated with stimulus perception, reward expectancy, movements related to the report and consumption of the reward, and increases in arousal. As highlighted in the introduction as well, in well-trained animals these signals are tightly correlated and hard to isolate. In our task, predictive representations need to be constructed from the different sensory inputs of future rewards and associated actions. The late report-related activity we report can therefore be interpreted in this context as the result of recurrent interactions between the visual cortex and downstream decision-making circuits in constructing such predictive representations going beyond a pure motor-related interpretation. The relevance of these predictive representations and what is predicted (reward or not) depends on task engagement and task context (whether visual stimuli are associated with rewarded actions and the existence of alternative auditory-informed actions). Alternative interpretations of the relevance of brain-wide multidimensional behavioral signals therefore include distributed associational learning, context-dependent sensory processing, or distributed motor command generation (Crochet et al., 2018; Engel et al., 2015; Kaplan and Zimmer, 2020; Wang et al., 2018).

Synthesis – How vision is affected by higher-order feedback, audition, perceptual outcome, and movement can in principle all be framed in the PC framework, where sensory

cortical hierarchies attempt to build a stable representation of the outside world. In our experiments, the central tenet that errors between predictions and sensory data are used to update internal representation was not specifically addressed. However, the machinery to support vision using signals from a wide variety of contextual sources is clearly there. It might be an interesting question for future research whether three types of signals (HVA feedback, auditory, and motor signals) engage the same predictive coding circuit in V1. In other words, under specific circumstances, contextual visual signals, auditory signals, and movements might predict the same visual input (for example oriented visual flow at some retinotopic location (Zmarz and Keller, 2016)) and this could be tested by intersecting methodological approaches to control visual predictions by sounds (Garner and Keller, 2022), visual context (Fiser et al., 2016; Keller et al., 2020a) and movement (Keller et al., 2012). A unifying predictive coding account of the visual cortical system would predict that the fundamental error computations are the same (irrespective of the source of the prediction) and thus would involve the same neuronal network, for example through the same negative and positive prediction error neurons (Jordan and Keller, 2020) and VIP disinhibitory circuits (Keller et al., 2020a).

Laminar organization of visual cortex modulation

Although we did not investigate error coding throughout our experiments, we did consistently sample activity across layers of V1 (Figure 1.5c). We can therefore compare the laminar organization of contextual modulation of V1 processing. According to the canonical microcircuit of the sensory cortex (Douglas and Martin, 2004; Harris and Shepherd, 2015), visual information from the dLGN mainly arrives in L4, propagates to L2/3 after which the strongest reciprocal connections are with L5. Layer 5 neurons constitute the main output of the cortex through local and long-range projections, with intracortical projections from L5a and subcortical projections from L5b. Layer 6 consists of pyramidal cells that primarily send feedback to the thalamus.

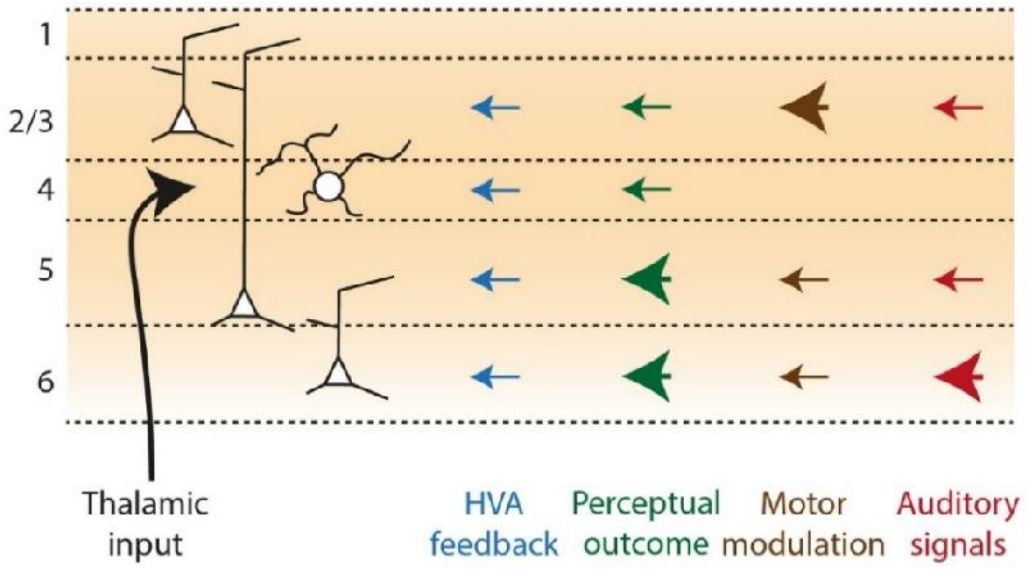


Figure 6.1 : Laminar organization of contextual modulation in V1 . *Left: Visual input arrives mostly in L4 from dLGN and is subsequently processed in L2/3 and L5/6. Right: Throughout the thesis, the laminar organization of contextual modulation was identified by pinpointing the spiking activity related to the various factors of interest. Arrows denote hotspots of somatic spiking, not the location of anatomical projections. Note that perceptual outcome and motor modulation are difficult to disentangle and overlap in time. Based on figures: 2.7d-f (HVA feedback), 3.2f, 3.4c (Perceptual outcome), 4.4a-b (Motor modulation and Auditory signals).*

We aligned the cortical depth of our V1 recordings using the laminar profile of the local field potential and the multi-unit activity (Fig. 2.7a; 3.2b; Supplementary Figure 4.6). Our multi-unit activity profile was consistent with previous reports with highest power in L5 due to its neurons with large somata and high discharge rates (Senzai et al., 2019). Also our current source density profiles were in line with earlier reports and were consistent with this propagation of visual information. The earliest current sink was visible in L4 and prominent sink-source pairs in L2/3 and L5/6 (Niell and Stryker, 2008; Schnabel et al., 2018; Senzai et al., 2019). This was also the case when analyzing at which timepoints neurons at which depth encoded visual information. The earliest visual encoding appeared in granular and supragranular layers and later in infragranular layers (Fig. 3.4 and Fig. 4.4).

We compared this spatiotemporal profile of visual coding to auditory coding and found that auditory-related activity was present in supragranular and especially infragranular layers (Fig. 4.4). Motor-related activity, on the other hand, was most prominent in supragranular layers, but also infragranular layers (Fig. 4.4). The report-related activity that we analyzed in Chapter 3, capturing visual hit/miss differences (including perceptual outcome as well as motor-related activity) was most prominent in supragranular and especially infragranular layers (Fig. 3.4), again avoiding the granular layer. We found that the inactivation of HVAs similarly affected all layers of V1 (Fig. 2.7).

Our results regarding the laminar specificity of contextual modulation can thus be summarized as in Figure 6.1 where the canonical visual information flow is represented alongside hotspots of modulations by sounds, perceptual outcome and motor activity, and feedback from HVAs. Even though anatomical connectivity and top-down modulation from higher visual areas (HVAs) are layer-specific (Harris et al., 2019; Keller et al., 2020b; Young et al., 2021), HVA feedback affected V1 activity throughout the layers. In general, however, contextual modulation was most prominent in extragranular layers. This is consistent with the idea that L4 is mostly driven by feedforward projections from dLGN, while infra- and supragranular layers contextualize and associate input through horizontal and feedback connections (Shen et al., 2020).

It is important to note that the depth of contextual modulation of spiking activity does not have to match anatomical projections. Indeed, nearly all neurons extend their dendrites to other layers with especially L5 pyramidal neurons sampling many layers and branching extensively in Layer 1. We found anatomical projections from auditory cortex to V1 in L1 and L5/6 (Fig. 4.14). This matched earlier reports of A1-V1 connectivity (Ibrahim et al., 2016; Rockland and Ojima, 2003), as well as that of feedback projections to V1 coming from higher visual areas (Young et al., 2021) and prefrontal cortex (Zhang et al., 2014) (Figure 6.2). The similarity of these laminar profiles suggests different cortical regions might recruit similar mechanisms to modulate V1 processing. For example, L1 neurons in V1 respond to movement signals as well as auditory noise bursts and sharpen visual

orientation tuning by a disinhibitory effect (Ibrahim et al., 2016; Mesik et al., 2019). A particular subset of L1 interneurons (NDNF-positive) has also been shown to modulate the gain of the cortical column upon fluctuations in arousal by acting on pyramidal apical dendrites and through other interneurons (Cohen-Kashi Malina et al., 2021). Also L6 has been shown to modulate the gain of visually-evoked responses either through feedback projections to the thalamus or through intracortical projections to other layers (Olsen et al., 2012). Furthermore, L6 acts as a site where head motion signals are integrated with visual signals (Vélez-Fort et al., 2018). Both L1 and L6 and perhaps to a lesser extent L2/3 and L5 can thus be seen as nodes through which convergent inputs from several brain areas can regulate the earliest steps of cortical visual processing (Harris and Shepherd, 2015; Schuman et al., 2021). Middle L4 is thus wrapped in a 'contextual sandwich' by superficial and deep layers.

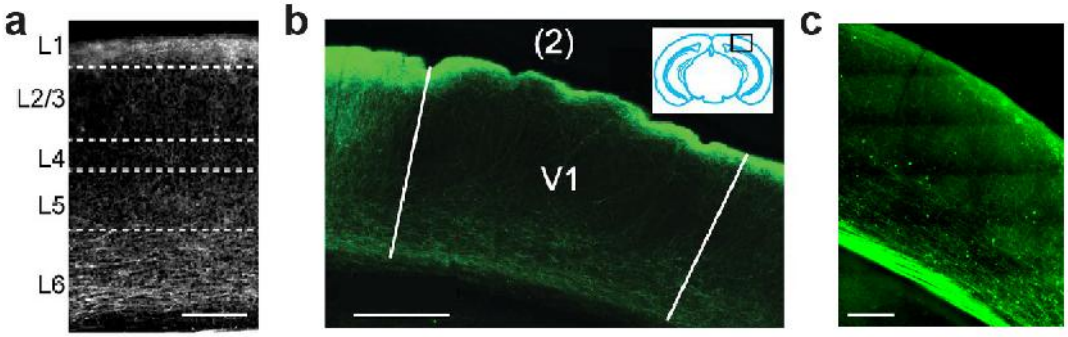


Figure 6.2: Projections from distinct sources of horizontal and top-down inputs to V1 have a similar laminar distribution of axons. ***a)*** *HVA feedback to V1. Anterogradely labeled feedback axons from area lateromedial to V1. Adapted from Young et al. (2021).* ***b)*** *Frontal top-down projections to V1. Fluorescence image in V1 of anterograde tracing from cingulate cortex. Adapted from Zhang et al. (2014).* ***c)*** *Auditory cortex projections to V1. Reproduced from Fig. 4.14 (this thesis). Scale bars are all 200 µm.*

It is worth emphasizing that feedforward and feedback (horizontal or top-down) communication are not temporally distinct or independent. In mouse visual cortex, feedforward and feedback projections are stronger on L5/6 neurons that project back to the source area (looped connectivity), suggesting a tight link between bottom-up and top-down information streams (Young et al., 2021). In primates, simultaneous recordings in V1, V2, and V4 have revealed that the integration of features across visual space depends on simultaneous and synergistic processing across areas (Chen et al., 2014, 2017; Liang et al., 2017).

Besides information about recording depth, our extracellular recordings allowed us to identify putative cell types based on the waveform (Figure 1.5b). We used this to verify our optogenetic strategy of driving narrow-spiking parvalbumin (PV)-expressing interneurons (Fig. 2.1 and 5.11). PV-interneurons are key regulators of rhythmic activity, most prominently gamma oscillations, and control the window of excitability of nearby pyramidal neurons (Bosman et al., 2014; Cardin et al., 2009; Sohal et al., 2009). When analyzing the responses of both broad- and narrow-spiking cell classes we found similar visually-evoked as well as report-related dynamics (Fig. 3.2) and similar effects of HVA inactivation (Fig. 2.7). Although we did not characterize tuning in detail, these results are

consistent with the findings that narrow-spiking cells are strongly recurrently connected to nearby excitatory neurons and play a main role in balancing excitation levels to prevent runaway excitation or suppression (Ahmadian et al., 2013; Ferguson and Cardin, 2020; Kim et al., 2018; Tsodyks et al., 1997; Xue et al., 2014).

It is clear that understanding the role of inhibitory neuron types is crucial to an understanding of cortical computation. For example, the differential circuit underlying the comparison of predicted and actual visual input in predictive coding is thought to involve specific inhibitory interneurons. For instance, the difference between excitatory inputs reflecting actual sensory input and inhibitory inputs (through VIP or SOM-expressing interneurons) reflecting movement-predicted input, could lead to neurons representing deviations from expected states (Attinger et al., 2017; Keller and Mrsic-Flogel, 2018). Although efforts have been made to differentiate more neuron types using extracellular recordings (Trainito et al., 2019; Vinck et al., 2015b), genetic approaches have proven incredibly powerful in approaching these issues.

Local versus distributed computation

Throughout our experiments, we recorded neuronal activity from a diverse set of cortical areas. Even though neural responses differed between areas, there was also substantial overlap (or redundancy). In Chapter 2 we recorded neuronal activity from V1 and two HVAs that were previously shown to differ most among all HVAs in their tuning to visual input: PM and AL (Andermann et al., 2011; Marshel et al., 2011). Although these areas show functional specialization in terms of preferred spatial and temporal frequencies of visual input, many characteristics also indicated substantial overlap. We recapitulated with extracellular recordings the earlier finding that PM and AL indeed show different tuning properties, while at the same time including neurons that prefer slow as well as fast stimuli (Fig. 2.1j). Moreover, silencing PM or AL produced qualitatively similar effects on V1 despite this tuning dissimilarity. Likewise, in multisensory trained animals (Chapters 3, 4, and 5), we found that neurons throughout the auditory, visual, and posterior parietal cortex (as well as medial prefrontal cortex; data not shown, manuscript in preparation), were modulated by vision, audition, and behavioral report (with the exception of visual coding in auditory cortex, which was strikingly absent). With variable coding distributed across recorded cortical areas, the question arises whether these areas are all doing something different or whether they collectively contribute to similar processes by processing information in a distributed fashion.

First, our findings are in line with comprehensive surveys of brain-wide activity that show that task-stimuli never evoke activity in single areas and that especially goal-directed action is associated with widespread activity changes (Allen et al., 2017b; Kauvar et al., 2020; Musall et al., 2019; Salkoff et al., 2020; Steinmetz et al., 2019). The fraction of neurons encoding these variables was different per region, suggesting functional gradients exist across the brain rather than homogeneous responses in individual areas (Tseng et al., 2022).

Most areas are connected in the mouse brain (Oh et al., 2014). Single neurons project routinely to multiple different target areas (Han et al., 2018) and these neurons form segregated subnetworks (Glickfeld et al., 2013a; Kim et al., 2018). The fact that there are very specific projections between areas and these form functional loops suggests that functional networks can exist that can form tight links, perhaps even functionally stronger

than within-area connectivity (Chen et al., 2013; Condylis et al., 2020; Mao et al., 2011; Petreanu et al., 2009; Young et al., 2021).
One interpretation that follows from the presence of such distributed signals is that computation is distributed (Christophel et al., 2017; Mejías and Wang, 2022; Thorpe, 1989). Instead of a set of local neurons performing computations, neurons across many areas form the functional substrate underlying a given behavior or task. An alternative interpretation is that areas and/or local circuits are strongly functionally specialized and that distributed signals provide the contextual information necessary for those local circuits to perform their relevant computations. This interpretation fits better with our reported dissociation between neural correlates and functional contributions to task performance. In spite of widespread visual, auditory, and report-related activity, visual detection was only affected by V1 inactivation (Chapter 3) and auditory detection only by auditory cortex inactivation (Chapter 4) and task-related activity in PPC was redundant to both visual and auditory task performance (Chapter 5). Although somewhat antagonistic, both interpretations suggest that information propagates through a densely connected network and it is not at all trivial to uncover which local signals or neuronal interactions within or between areas are involved in the relevant computations.

Task-dependent network reconfigurations

Several conclusions in this thesis were the result of monitoring and perturbing the same circuitry while varying behavioral relevance and task requirements. First, the causal involvement of V1 was extended in the multisensory task context relative to unisensory trained mice (Chapter 3) and this conclusion would not have been reached without either causal interventions or varying task context. Second, varying behavioral relevance across mice led to different amounts of instructed orofacial movements (licking) for each modality, which allowed to dissociate motor and auditory origins of sound-evoked activity across multisensory task context (Chapter 4). Third, this approach revealed that the response magnitude of PPC neurons to visual and auditory stimuli during audiovisual change detection was enhanced in trained versus naive mice, even though this was not associated with an acquired causal role in the task (Chapter 5).

These results emphasize the powerful insights that might be obtained by contrasting neural circuit functioning between contexts that differ in task demands and associated required strategy. A study by Pinto *et al.* (2019) similarly reported that increasing task demands by introducing evidence accumulation or working memory in a virtual navigation task (keeping roughly similar visual cues) increased the importance and task-related activity in various dorsal cortical areas such as parietal cortex. This approach might be further guided and constrained by computational theory and statistical models of task-performing neural circuits (Linderman and Gershman, 2017; Maheswaranathan et al., 2018). For example, modeling of recurrent neural networks trained to produce required behavioral outputs as in a perceptual decision making task allows *in silico* perturbations to test the causal role of specific parts of the circuitry in such a task (Pinto et al., 2019; Yang et al., 2019).

Nevertheless, we only varied behavioral context between animals, and not within. It would be interesting for future experiments to tease out whether task-dependent reconfigurations are the result of extensive behavioral training or can depend on the immediate task context. In a detection task in which the likelihood of tactile and visual stimuli was varied across sessions, rats responded faster to the same stimuli in high-likelihood sessions, suggesting reprioritization across modalities (Lee et al., 2016). This was shown to be even possible on

a trial-by-trial basis if the relevant modality was cued (Wimmer et al., 2015). In Chapter 3 we report the effect of adding another relevant modality on the duration of involvement of V1 between cohorts of mice. Possibly V1 involvement could dynamically vary based on whether - in a given trial - also other modalities need to be monitored. Alternatively, extended training on multiple modalities shapes the way V1 and the rest of the circuitry solve the task, and training history determines circuit implementation (Arlt et al., 2021). A further avenue is to consider potential variability in strategies between animals (Waskom et al., 2019). In an auditory and tactile short-term memory task, mice displayed different strategies to solve the task. These strategies were associated with different patterns of neural activity as well as causal involvement across dorsal cortical areas (Gallero-Salas et al., 2021; Gilad et al., 2018).

Similarly, we employed acute recordings and sampled different neurons in each session. It would be interesting to track the course of neuronal activity across days to identify the plastic changes underlying the differential network configurations underlying our task-dependent differences. Tracking the activity of the same neuronal populations either with implanted electrodes or with chronic calcium imaging (Andermann et al., 2010; Kim et al., 2016), has shown that during learning of a perceptual task the selectivity of neurons changes (Poort et al., 2015) and noise correlations between neurons decrease (Najafi et al., 2020; Ni et al., 2018), both resulting in improved representations of rewarded stimuli (Goltstein et al., 2013, 2018).

Circuit causally underlying audiovisual change detection

We reported a stark contrast between the presence of distributed sensory and report-related information and direct functional contribution to auditory and visual change detection. We found that optogenetic inactivations of V1 impaired visual change detection and muscimol inactivations of auditory cortex impaired auditory change detection (Chapters 3 and 4). Classically, higher-order regions such as the parietal cortex have been considered the main regions of interest to perform multisensory processing and decision making (Hanks et al., 2015; Raposo et al., 2014; Stein and Stanford, 2008; Wallace et al., 2004). In our experiments, this is a region that is surprisingly not involved in task performance, or at least to be redundant relative to other circuits. Can we pinpoint the circuitry required to solve the task?

Auditory and visual signals may bypass PPC and propagate from primary sensory areas to more frontal areas or other areas directly, instructing behavioral output. Alternatively, these signals could propagate via more lateral higher-order areas in between the primary visual and auditory cortex, such as the anterolateral area (AL) (Meijer et al., 2020; Oh et al., 2014; Wang et al., 2011; Zingg et al., 2014). It is further possible that fast network reconfigurations compensate for the local loss of function and recruit redundant cortical pathways. However, it is even more likely that subcortical areas, not emphasized in this corticocentric thesis, play an important role. First, multisensory interactions are abundant in subcortical circuits such as the thalamus (Allen et al., 2017a; Bieler et al., 2018) and superior colliculus (Meredith and Stein, 1986; Stein and Stanford, 2008). In line with this, optogenetic silencing of V1 did not affect the detection of high-contrast contralateral visual stimuli during visuotactile side detection (Fig. 3.6), suggesting subcortical visual circuits, such as the superior colliculus, might be sufficient to perform visual detection and orienting tasks (Krauzlis et al., 2013, 2018; Wang et al., 2020). Incomplete V1 inactivation might also explain these results. Cortical projections that bypass parietofrontal areas and go

directly to the relevant parts of the striatum can also drive sensory-based decisions (Znamenskiy and Zador, 2013). Recent technological developments have generated a scaling up in the number of simultaneously probed neurons and areas that is promising to allow a less restrictive focus on a subset of areas and in principle design less biased experiments. These approaches have pointed to relatively unknown subcortical areas, such as the *zona incerta,* that display surprising selectivity for choice, but not action, during visual decision making (Steinmetz et al., 2019).

'Grounding' neural activity

I would also like to explicate some of the implicit frameworks that underlie much of current systems neuroscience, as well as the studies and analyses presented in this thesis. As is routinely done in sensory neuroscience experiments, we varied sensory variables along prespecified dimensions (grating orientation or tone frequency) and search in the brain for those activity patterns that correlate with chosen variables. We average the spiking rate in an experimenter-defined time window across subsets of trials throughout a session. Neurons that correlate with these experimental variables are said to 'encode' these properties. However, the brain has no access to any of these variables: it merely processes spikes. This issue has often been expressed by various researchers (Buzsáki, 2019; Gregory, 1980; Helmholtz, 1867; Pennartz, 2015, 2009), and quickly forgotten or disregarded upon analyzing neural data.

However, for the brain to know that activity increments or decrements relate to specific external variables, additional context or a cipher is necessary. In many studies, the problem that some variable that is encoded must also be deciphered is relegated elsewhere and routinely solved by postulating the existence of downstream 'decoding' neurons; this poses the question of whether there are separate encoding and decoding neurons or distinct decoding parts of the brain (Brette, 2019). This problem is more generally called the 'grounding problem': how do neural circuits figure out what their activity relates to in the external world (Harnad, 1990). Several researchers have offered different solutions, such as differentiating/grounding signals through the relational properties of their input statistics within, but importantly also across, the sensory modalities (Pennartz, 2015, 2009). Others emphasize the need for action-based feedback interactions and the need to ground sensory inputs by exploring movement-induced sensory consequences (Buzsáki, 2019; Chemero, 2009; Clark, 2013; Friston, 2010).

Importantly, these viewpoints emphasize the need to understand how neural activity is used and makes sense *within* the brain and what is available locally. For example, in Chapter 3 we reported auditory responses in V1 neurons: how do these neurons 'know' that they are representing auditory features? In Pennartz *et al.* (2009), this is posited as a distinct problem at the level of the different senses, i.e. the modality identification problem. How come visually-induced spiking activity has a qualitatively different character from auditory-induced activity, even though both consist of a pattern of spikes distributed across neurons? At the level of sensory modalities, one might propose that crossmodal interactions between the senses are necessary to differentiate and calibrate the modally distinct activity patterns (Pennartz, 2009). This identification problem holds however for all features and submodalities: what grounds the activity of V1 neurons as relating to certain oriented contrasts in the external world and not others? To avoid regression into needing 'identification machinery' for each neuron, one natural avenue is to search for a higher-order mechanism that arises through the collective interaction of distributed

neurons (Pennartz, 2015). Others pose that the causal structure between elements (neurons) determines the qualitative nature of the experience (Tononi et al., 2016).

One analytical approach to come closer to this problem is to take the brain's perspective when analyzing data. Making sense of physiological data without any knowledge of experimental variables appears as a profound analytical challenge, but that is the situation neural circuits are in, trapped inside a skull. For example, we might revisit and extend our interpretation of the results in Chapter 5 as follows. In this chapter, we observed single neuron and population 'coding' of visual, auditory, and hit/miss variables (using the experimental knowledge about the trials in the task). However, optogenetic inactivation did not affect task performance. We concluded that it is important to search for *those* activity patterns that directly and causally inform perceptual decision making (cf. Panzeri et al., 2017; Ritchie et al., 2016). A further specification is to search for activity that is task-informative and *makes sense* to the rest of the neural circuit given its current state (and not only to the experimenter). What are the learning rules or circuit motifs that allow activity to acquire meaning intrinsic to the system? This analytical exercise seems more tractable when taking brain development into account and with more complete circuit data than a subset of V1 neurons. Investigating these questions will rely on (i) developing methods and experimental approaches that take the brain's perspective and thus have access to local and ongoing neuronal dynamics, without knowledge of external experimental variables (cf. Mayner et al., 2022), as well as (ii) developing causal perturbations that interrogate the circuit in a closed-loop manner and induce physiologically meaningful activity patterns (Jazayeri and Afraz, 2017). This approach might help in starting to understand how the brain makes inferences and constructs representations that are meaningful within the brain.

Methodological limitations

Despite the power of head-fixed experiments with controlled sensory stimuli, we should note limitations of this approach. To study the mechanistic underpinnings of how responses of single neurons in visual and auditory cortex encode sensory features and how this processing is flexibly modulated, we used simple gratings and static harmonic tones. These low-dimensional stimuli are optimized to elicit predictable and well-described activity in their respective cortices as neurons in the visual cortex are tuned to oriented edges (Hubel and Wiesel, 1979; Marr and Hildreth, 1980), while auditory cortical neurons are tuned to frequency-specific on- and offsets (Chong et al., 2020; Issa et al., 2014; Scholl et al., 2010). Especially in combination with head-fixed conditions with two-alternative response options, this allowed us to tightly control the sensory inputs and dissociate feature-specific processing (stimulus A or B), from signaling the stimulus presence and modality (visual, auditory, or no stimulus) and behavioral variables (Chapter 3, 4, and 5). Furthermore, this made the task cortex-dependent as V1 was necessary for visual performance (Chapter 3) and AC for auditory performance (Chapter 4).

Despite these clear advantages, combining low-dimensional cortex-optimized stimuli with low-dimensional behavioral output under head-fixed conditions also raises problems. This concern has been voiced in various ways (Juavinett et al., 2018; Kaplan and Zimmer, 2020; Krakauer et al., 2017), but I would like to discuss two points particularly relevant to the current thesis. First, head-fixation is quite an unnatural condition. Under head-fixed conditions, contractions of the head and body muscles cease to evoke their normal expected changes in sensory inputs, disrupting the continuity of normal sensorimotor loops. We

found striking neural activity patterns that were to a considerable extent explained by motor variables in V1 as well as PPC (Chapters 3, 4, and 5). A fuller interpretation of our results will benefit from further research on the role of these widespread motor correlates (see also above); what is their role in freely moving contexts and consequently, and what are the functional consequences of partially discontinued sensorimotor functioning under head fixation (Parker et al., 2020)? The importance of sensorimotor interactions with the world is evident in naturalistic behavioral paradigms, either freely moving or through head-fixed virtual reality, and these are associated with much faster learning rates compared to two-alternative stimulus-response paradigms (Havenith et al., 2018; Rosenberg et al., 2021). Developments in the application of machine learning techniques to quantify animal behavior in videos will aid in a more detailed characterization of animal movements with high precision and a better understanding of the animal's strategy (Kabra et al., 2013; Mathis et al., 2018; Robie et al., 2017).

Second, in the multisensory version of our behavioral task (MST), we required animals to discriminate the modalities and respond oppositely to visual and auditory changes. This specific perceptual task – the grouping of non-naturalistic stimuli according to their modality – does not directly map onto ecological behaviors (e.g. always take a left at an intersection if you see something, but go right if you hear something). Therefore, we should take extra caution in how such settings lead to insights into brain functioning that generalize to more ecological settings and humans. Of course, humans can readily report the modality of presented stimuli and this seems inextricably linked to the perception of these stimuli (Lehman, 1982; Overgaard et al., 2013). It is only after recall that they sometimes turn out to have forgotten if they heard or saw something (Lehman, 1982). Overall, however, this behavioral paradigm is an important step towards a broader view of multisensory processing beyond cue integration (Meijer et al., 2019).

Translation to humans

How do these results improve our understanding of all brains, including those of humans? First, it is important to question to what extent primary visual cortex in mice corresponds to primate V1. We report striking non-visual signals already in primary visual cortex. Whether this is specific to mice or also holds for primate V1 is unclear. Classic work in primates suggests that auditory signals do indeed affect activity in early visual cortices, but report only modulatory effects, rather than evoking responses by themselves (Lakatos et al., 2009; Molholm et al., 2002; Petro et al., 2017; Senkowski et al., 2011; Wang et al., 2008) - but see (Brang et al., 2022). We further report motor-related activity in V1. It is known that the primate visual cortex is modulated by eye movements (Supèr et al., 2004; Wurtz, 1968), but whether the same degree of sensorimotor integration is present with body movements affecting a large fraction of V1 neurons, is yet unclear (Busse et al., 2017; Froudarakis et al., 2019).

However, similar to mice, humans use vision in very flexible ways and it is very clear that human vision ought to be and is strongly influenced by context. Human vision is similarly affected by the discussed contextual factors as evidenced by the many multisensory illusions (Mcgurk and Macdonald, 1976; Sekuler et al., 1997; Shams et al., 2000), attentional task effects (Simons and Chabris, 1999) and movement effects (Simons and Levin, 1997). This flexible use of vision depends on cortical circuits in occipital and parietal cortex to integrate contextual information with incoming visual information to guide action.

Conclusion

Overall, the results presented in this thesis suggest that factors beyond immediate visual input are fundamental to basic features of visual processing. Feedback projections shape basic visual response properties to facilitate robust encoding and, besides visual recurrent signals, also sounds, movements and arousal strongly affect V1 activity. These signals are dependent on the multisensory task context and should be carefully disentangled as they constitute separate components present alongside visual inputs. Furthermore, visual processing is not a static process. Even the involvement of processing at the first cortical stage in the context of plain stimulus detection depends on the task context, as instantiated via other modalities. The fact that optogenetic silencing of V1 specifically affected visual but not auditory perception defends the idea that V1 is primarily involved in visual processing, but otherwise *primary visual* cortex seems hardly visual alone. It just needs some context.

Bibiography

Aarts, E., Verhage, M., Veenvliet, J.V., Dolan, C.V., and van der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. Nat Neurosci *17*, 491–496. https://doi.org/10.1038/nn.3648.

Abbott, L.F., and Dayan, P. (1999). The Effect of Correlated Variability on the Accuracy of a Population Code. Neural Computation *11*, 91–101. https://doi.org/10.1162/089976699300016827.

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z.J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. Nature *490*, 226–231. https://doi.org/10.1038/nature11526.

Ahmadian, Y., Rubin, D.B., and Miller, K.D. (2013). Analysis of the Stabilized Supralinear Network. Neural Computation *25*, 1994–2037. https://doi.org/10.1162/NECO_a_00472.

Akrami, A., Kopec, C.D., Diamond, M.E., and Brody, C.D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. Nature *554*, 368–372. https://doi.org/10.1038/nature25510.

Alkire, M.T., Hudetz, A.G., and Tononi, G. (2008). Consciousness and anesthesia. Science *322*, 876–880. https://doi.org/10.1126/science.1149213.

Allen, A.E., Procyk, C.A., Brown, T.M., and Lucas, R.J. (2017a). Convergence of visual and whisker responses in the primary somatosensory thalamus (ventral posterior medial region) of the mouse. J Physiol *595*, 865–881. https://doi.org/10.1113/JP272791.

Allen, W.E., Kauvar, I.V., Chen, M.Z., Richman, E.B., Yang, S.J., Chan, K., Gradinaru, V., Deverman, B.E., Luo, L., and Deisseroth, K. (2017b). Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. Neuron *94*, 891-907.e6. https://doi.org/10.1016/j.neuron.2017.04.017.

Allman, J., Miezin, F., and McGuinness, E. (1985). Stimulus Specific Responses from Beyond the Classical Receptive Field: Neurophysiological Mechanisms for Local-Global Comparisons in Visual Neurons. Annual Review of Neuroscience *8*, 407–430. https://doi.org/10.1146/annurev.ne.08.030185.002203.

Andermann, M.L., Kerlin, A.M., and Reid, C. (2010). Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. Front. Cell. Neurosci. *4*. https://doi.org/10.3389/fncel.2010.00003.

Andermann, M.L., Kerlin, A.M., Roumis, D.K., Glickfeld, L.L., and Reid, R.C. (2011). Functional specialization of mouse higher visual cortical areas. Neuron *72*, 1025–1039. https://doi.org/10.1016/j.neuron.2011.11.013.

Andersen, R.A., and Cui, H. (2009). Intention, Action Planning, and Decision Making in Parietal-Frontal Circuits. Neuron *63*, 568–583. https://doi.org/10.1016/j.neuron.2009.08.028.

Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses. Science *273*, 1868–1871. https://doi.org/10.1126/science.273.5283.1868.

Arlt, C., Barroso-Luque, R., Kira, S., Bruno, C.A., Xia, N., Chettih, S.N., Soares, S., Pettit, N.L., and Harvey, C.D. (2021). Cognitive experience alters cortical involvement in navigation decisions. BioRxiv 2021.12.10.472106. https://doi.org/10.1101/2021.12.10.472106.

Attinger, A., Wang, B., and Keller, G.B. (2017). Visuomotor Coupling Shapes the Functional Development of Mouse Visual Cortex. Cell *169*, 1291-1302.e14. https://doi.org/10.1016/j.cell.2017.05.023.

Attneave, F. (1954). Some informational aspects of visual perception. Psychological Review *61*, 183–193. https://doi.org/10.1037/h0054663.

Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nature Reviews Neuroscience *7*, 358–366. https://doi.org/10.1038/nrn1888.

Azouz, R., and Gray, C.M. (1999). Cellular Mechanisms Contributing to Response Variability of Cortical Neurons In Vivo. J. Neurosci. *19*, 2209–2223. .

Baldwin, M.K.L., Balaram, P., and Kaas, J.H. (2017). The evolution and functions of nuclei of the visual pulvinar in primates. Journal of Comparative Neurology *525*, 3207–3226. https://doi.org/10.1002/cne.24272.

Bao, S., Chan, V.T., and Merzenich, M.M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. Nature *412*, 79–83. https://doi.org/10.1038/35083586.

Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. In Sensory Communication, W. Rosenblith, ed. (MIT Press), pp. 217–234.

Barthas, F., and Kwan, A.C. (2016). Secondary Motor Cortex: Where "Sensory" Meets "Motor" in the Rodent Frontal Cortex. Trends Neurosci. https://doi.org/10.1016/j.tins.2016.11.006.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., De Weerd, P., Kennedy, H., and Fries, P. (2015). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. Neuron *85*, 390–401. https://doi.org/10.1016/j.neuron.2014.12.018.

Beaman, C.B., Eagleman, S.L., and Dragoi, V. (2017). Sensory coding accuracy and perceptual performance are improved during the desynchronized cortical state. Nature Communications *8*, 1308. https://doi.org/10.1038/s41467-017-01030-4.

Bender, M.B., and Teuber, H.L. (1947). Spatial organization of visual perception following injury to the brain. Arch Neurol Psychiatry *58*, 721; passim. https://doi.org/10.1001/archneurpsyc.1947.02300350071007.

Bennett, C., Arroyo, S., and Hestrin, S. (2013). Subthreshold Mechanisms Underlying State-Dependent Modulation of Visual Responses. Neuron *80*, 350–357. https://doi.org/10.1016/j.neuron.2013.08.007.

Bennett, C., Gale, S.D., Garrett, M.E., Newton, M.L., Callaway, E.M., Murphy, G.J., and Olsen, S.R. (2019). Higher-Order Thalamic Circuits Channel Parallel Streams of Visual Information in Mice. Neuron *102*, 477-492.e5. https://doi.org/10.1016/j.neuron.2019.02.010.

Berezovskii, V.K., Nassi, J.J., and Born, R.T. (2011). Segregation of feedforward and feedback projections in mouse visual cortex. J. Comp. Neurol. *519*, 3672–3683. https://doi.org/10.1002/cne.22675.

Bieler, M., Xu, X., Marquardt, A., and Hanganu-Opatz, I.L. (2018). Multisensory integration in rodent tactile but not visual thalamus. Scientific Reports *8*, 15684. https://doi.org/10.1038/s41598-018-33815-y.

Bimbard, C., Sit, T.P., Lebedeva, A., Harris, K.D., and Carandini, M. (2021). Behavioral origin of sound-evoked activity in visual cortex. BioRxiv 2021.07.01.450721. https://doi.org/10.1101/2021.07.01.450721.

Bisley, J.W., and Goldberg, M.E. (2010). Attention, Intention, and Priority in the Parietal Lobe. Annual Review of Neuroscience *33*, 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823.

Bizley, J.K., Nodal, F.R., Bajo, V.M., Nelken, I., and King, A.J. (2007). Physiological and Anatomical Evidence for Multisensory Interactions in Auditory Cortex. Cereb. Cortex *17*, 2172–2189. https://doi.org/10.1093/cercor/bhl128.

Blake, D.T., Heiser, M.A., Caywood, M., and Merzenich, M.M. (2006). Experience-Dependent Adult Cortical Plasticity Requires Cognitive Association between Sensation and Reward. Neuron *52*, 371–381. https://doi.org/10.1016/j.neuron.2006.08.009.

Blot, A., Roth, M.M., Gasler, I., Javadzadeh, M., Imhof, F., and Hofer, S.B. (2021). Visual intracortical and transthalamic pathways carry distinct information to cortical areas. Neuron *109*, 1996-2008.e6. https://doi.org/10.1016/j.neuron.2021.04.017.

Boehler, C.N., Tsotsos, J.K., Schoenfeld, M.A., Heinze, H.-J., and Hopf, J.-M. (2009). The center-surround profile of the focus of attention arises from recurrent processing in visual cortex. Cereb Cortex *19*, 982–991. https://doi.org/10.1093/cercor/bhn139.

Boly, M., Garrido, M.I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S., and Friston, K. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. Science *332*, 858–862. https://doi.org/10.1126/science.1202043.

Bos, J.J., Vinck, M., van Mourik-Donga, L.A., Jackson, J.C., Witter, M.P., and Pennartz, C.M.A. (2017). Perirhinal firing patterns are sustained across large spatial segments of the task environment. Nature Communications *8*, 15602. https://doi.org/10.1038/ncomms15602.

Bos, J.J., Vinck, M., Marchesi, P., Keestra, A., van Mourik-Donga, L.A., Jackson, J.C., Verschure, P.F.M.J., and Pennartz, C.M.A. (2019). Multiplexing of Information about Self and Others in Hippocampal Ensembles. Cell Reports *29*, 3859-3871.e6. https://doi.org/10.1016/j.celrep.2019.11.057.

Bosman, C.A., Lansink, C.S., and Pennartz, C.M.A. (2014). Functions of gamma-band synchronization in cognition: from single circuits to functional diversity across cortical and subcortical systems. European Journal of Neuroscience *39*, 1982–1999. https://doi.org/10.1111/ejn.12606.

Bouvier, G., Senzai, Y., and Scanziani, M. (2020). Head Movements Control the Activity of Primary Visual Cortex in a Luminance-Dependent Manner. Neuron *108*, 500-511.e5. https://doi.org/10.1016/j.neuron.2020.07.004.

Brang, D., Plass, J., Sherman, A., Stacey, W.C., Wasade, V.S., Grabowecky, M., Ahn, E., Towle, V.L., Tao, J.X., Wu, S., et al. (2022). Visual cortex responds to sound onset and offset during passive listening. Journal of Neurophysiology https://doi.org/10.1152/jn.00164.2021.

Brette, R. (2019). Is coding a relevant metaphor for the brain? Behavioral and Brain Sciences *42*. https://doi.org/10.1017/S0140525X19000049.

Brody, C.D., and Hanks, T.D. (2016). Neural underpinnings of the evidence accumulator. Current Opinion in Neurobiology *37*, 149–157. https://doi.org/10.1016/j.conb.2016.01.003.

Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. Science *340*, 95–98. https://doi.org/10.1126/science.1233912.

Budinger, E., and Scheich, H. (2009). Anatomical connections suitable for the direct processing of neuronal information of different modalities via the rodent primary auditory cortex. Hearing Research *258*, 16–27. https://doi.org/10.1016/j.heares.2009.04.021.

Budinger, E., Heil, P., Hess, A., and Scheich, H. (2006). Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. Neuroscience *143*, 1065–1083. https://doi.org/10.1016/j.neuroscience.2006.08.035.

Burgess, C.P., Lak, A., Steinmetz, N.A., Zatka-Haas, P., Reddy, C.B., Jacobs, E.A.K., Linden, J.F., Paton, J.J., Ranson, A., Schröder, S., et al. (2017). High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. Cell Reports *20*, 2513–2524. https://doi.org/10.1016/j.celrep.2017.08.047.

Burkhalter, A. (2016). The Network for Intracortical Communication in Mouse Visual Cortex. In Micro-, Meso- and Macro-Connectomics of the Brain, H. Kennedy, D.C.V. Essen, and Y. Christen, eds. (Springer International Publishing), pp. 31–43.

Burr, D., Banks, M.S., and Morrone, M.C. (2009). Auditory dominance over vision in the perception of interval duration. Exp Brain Res *198*, 49. https://doi.org/10.1007/s00221-009-1933-z.

Busse, L., Ayaz, A., Dhruv, N.T., Katzner, S., Saleem, A.B., Scholvinck, M.L., Zaharia, A.D., and Carandini, M. (2011). The Detection of Visual Contrast in the Behaving Mouse. Journal of Neuroscience *31*, 11351–11361. https://doi.org/10.1523/JNEUROSCI.6689-10.2011.

Busse, L., Cardin, J.A., Chiappe, M.E., Halassa, M.M., McGinley, M.J., Yamashita, T., and Saleem, A.B. (2017). Sensation during Active Behaviors. J. Neurosci. *37*, 10826–10834. https://doi.org/10.1523/JNEUROSCI.1828-17.2017.

Buzsáki, G. (2019). The Brain from Inside Out (Oxford University Press).

Cai, L., Wu, B., and Ji, S. (2018). Neuronal Activities in the Mouse Visual Cortex Predict Patterns of Sensory Stimuli. Neuroinformatics *16*, 473–488. https://doi.org/10.1007/s12021-018-9357-1.

Campi, K.L., Bales, K.L., Grunewald, R., and Krubitzer, L. (2010). Connections of Auditory and Visual Cortex in the Prairie Vole (Microtus ochrogaster): Evidence for Multisensory Processing in Primary Sensory Areas. Cereb. Cortex *20*, 89–108. https://doi.org/10.1093/cercor/bhp082.

Cappe, C., and Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. European Journal of Neuroscience *22*, 2886–2902. https://doi.org/10.1111/j.1460-9568.2005.04462.x.

Carandini, M., and Churchland, A.K. (2013). Probing perceptual decisions in rodents. Nat Neurosci *16*, 824–831. https://doi.org/10.1038/nn.3410.

Cardin, J.A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.-H., and Moore, C.I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature *459*, 663–667. https://doi.org/10.1038/nature08002.

Ceballo, S., Piwkowska, Z., Bourg, J., Daret, A., and Bathellier, B. (2019). Targeted Cortical Manipulation of Auditory Perception. Neuron *104*, 1168-1179.e5. https://doi.org/10.1016/j.neuron.2019.09.043.

Chemero, A. (2009). Radical embodied cognitive science (Cambridge, Mass: MIT Press).

Chen, J.L., Carta, S., Soldado-Magraner, J., Schneider, B.L., and Helmchen, F. (2013). Behaviour-dependent recruitment of long-range projection neurons in somatosensory cortex. Nature *499*, 336–340. https://doi.org/10.1038/nature12236.

Chen, J.L., Voigt, F.F., Javadzadeh, M., Krueppel, R., and Helmchen, F. (2016). Long-range population dynamics of anatomically defined neocortical networks. ELife *5*, e14679. https://doi.org/10.7554/eLife.14679.

Chen, M., Yan, Y., Gong, X., Gilbert, C.D., Liang, H., and Li, W. (2014). Incremental Integration of Global Contours through Interplay between Visual Cortical Areas. Neuron *82*, 682–694. https://doi.org/10.1016/j.neuron.2014.03.023.

Chen, R., Wang, F., Liang, H., and Li, W. (2017). Synergistic Processing of Visual Contours across Cortical Layers in V1 and V2. Neuron *96*, 1388-1402.e4. https://doi.org/10.1016/j.neuron.2017.11.004.

Choi, I., Lee, J.-Y., and Lee, S.-H. (2018). Bottom-up and top-down modulation of multisensory integration. Current Opinion in Neurobiology *52*, 115–122. https://doi.org/10.1016/j.conb.2018.05.002.

Chong, K.K., Anandakumar, D.B., Dunlap, A.G., Kacsoh, D.B., and Liu, R.C. (2020). Experience-Dependent Coding of Time-Dependent Frequency Trajectories by Off Responses in Secondary Auditory Cortex. J. Neurosci. *40*, 4469–4482. https://doi.org/10.1523/JNEUROSCI.2665-19.2020.

Chou, X., Fang, Q., Yan, L., Zhong, W., Peng, B., Li, H., Wei, J., Tao, H.W., and Zhang, L.I. (2020). Contextual and cross-modality modulation of auditory cortical processing through pulvinar mediated suppression. ELife *9*, e54157. https://doi.org/10.7554/eLife.54157.

Christophel, T.B., Klink, P.C., Spitzer, B., Roelfsema, P.R., and Haynes, J.-D. (2017). The Distributed Nature of Working Memory. Trends in Cognitive Sciences *21*, 111–124. https://doi.org/10.1016/j.tics.2016.12.007.

Chubykin, A.A., Roach, E.B., Bear, M.F., and Shuler, M.G.H. (2013). A Cholinergic Mechanism for Reward Timing within Primary Visual Cortex. Neuron *77*, 723–735. https://doi.org/10.1016/j.neuron.2012.12.039.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences *36*, 181–204. https://doi.org/10.1017/S0140525X12000477.

Coen, P., Sit, T.P.H., Wells, M.J., Carandini, M., and Harris, K.D. (2021). Mouse frontal cortex mediates additive multisensory decisions. BioRxiv 2021.04.26.441250. https://doi.org/10.1101/2021.04.26.441250.

Cohen, Y.E. (2009). Multimodal activity in the parietal cortex. Hear Res *258*, 100–105. https://doi.org/10.1016/j.heares.2009.01.011.

Cohen, M.R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. Nat Neurosci *14*, 811–819. https://doi.org/10.1038/nn.2842.

Cohen, M.R., and Maunsell, J.H.R. (2009). Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci *12*, 1594–1600. https://doi.org/10.1038/nn.2439.

Cohen-Kashi Malina, K., Tsivourakis, E., Kushinsky, D., Apelblat, D., Shtiglitz, S., Zohar, E., Sokoletsky, M., Tasaka, G., Mizrahi, A., Lampl, I., et al. (2021). NDNF interneurons in layer 1 gain-modulate whole cortical columns according to an animal's behavioral state. Neuron *109*, 2150-2164.e5. https://doi.org/10.1016/j.neuron.2021.05.001.

Condylis, C., Lowet, E., Ni, J., Bistrong, K., Ouellette, T., Josephs, N., and Chen, J.L. (2020). Context-Dependent Sensory Processing across Primary and Secondary Somatosensory Cortex. Neuron *106*, 515-525.e5. https://doi.org/10.1016/j.neuron.2020.02.004.

Cossell, L., Iacaruso, M.F., Muir, D.R., Houlton, R., Sader, E.N., Ko, H., Hofer, S.B., and Mrsic-Flogel, T.D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. Nature *518*, 399–403. https://doi.org/10.1038/nature14182.

Crapse, T.B., and Sommer, M.A. (2008). Corollary discharge across the animal kingdom. Nat Rev Neurosci *9*, 587–600. https://doi.org/10.1038/nrn2457.

Crochet, S., and Petersen, C.C.H. (2006). Correlating whisker behavior with membrane potential in barrel cortex of awake mice. Nature Neuroscience *9*, 608–610. https://doi.org/10.1038/nn1690.

Crochet, S., Lee, S.-H., and Petersen, C.C.H. (2018). Neural Circuits for Goal-Directed Sensorimotor Transformations. Trends in Neurosciences https://doi.org/10.1016/j.tins.2018.08.011.

Cui, H., and Andersen, R.A. (2007). Posterior Parietal Cortex Encodes Autonomously Selected Motor Plans. Neuron *56*, 552–559. https://doi.org/10.1016/j.neuron.2007.09.031.

Cul, A.D., Baillet, S., and Dehaene, S. (2007). Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness. PLOS Biology *5*, e260. https://doi.org/10.1371/journal.pbio.0050260.

Cullen, K.E. (2019). Vestibular processing during natural self-motion: implications for perception and action. Nat Rev Neurosci *20*, 346–363. https://doi.org/10.1038/s41583-019-0153-1.

Dadarlat, M.C., and Stryker, M.P. (2017). Locomotion Enhances Neural Encoding of Visual Stimuli in Mouse V1. J. Neurosci. *37*, 3764–3775. https://doi.org/10.1523/JNEUROSCI.2728-16.2017.

De Franceschi, G., and Barkat, T.R. (2021). Task-induced modulations of neuronal activity along the auditory pathway. Cell Reports *37*, 110115. https://doi.org/10.1016/j.celrep.2021.110115.

De Meo, R., Murray, M.M., Clarke, S., and Matusz, P.J. (2015). Top-down control and early multisensory processes: chicken vs. egg. Front Integr Neurosci *9*. https://doi.org/10.3389/fnint.2015.00017.

Dehaene, S., and Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. Neuron *70*, 200–227. https://doi.org/10.1016/j.neuron.2011.03.018.

Deneux, T., Harrell, E.R., Kempf, A., Ceballo, S., Filipchuk, A., and Bathellier, B. (2019). Context-dependent signaling of coincident auditory and visual events in primary visual cortex. ELife *8*, e44006. https://doi.org/10.7554/eLife.44006.

Denny-Brown, D., Meyer, J.S., and Horenstein, S. (1952). The significance of perceptual rivalry resulting from parietal lesion. Brain *75*, 433–471. https://doi.org/10.1093/brain/75.4.432.

Desimone, R., and Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. Annu. Rev. Neurosci. *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205.

Deweese, M.R., and Zador, A.M. (2004). Shared and Private Variability in the Auditory Cortex. Journal of Neurophysiology *92*, 1840–1855. https://doi.org/10.1152/jn.00197.2004.

DeWeese, M.R., Wehr, M., and Zador, A.M. (2003). Binary Spiking in Auditory Cortex. J. Neurosci. *23*, 7940–7949. https://doi.org/10.1523/JNEUROSCI.23-21-07940.2003.

Dias, R., Robbins, T.W., and Roberts, A.C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. Nature *380*, 69–72. https://doi.org/10.1038/380069a0.

DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? Neuron *73*, 415–434. https://doi.org/10.1016/j.neuron.2012.01.010.

Douglas, R.J., and Martin, K.A.C. (2004). Neuronal Circuits of the Neocortex. Annual Review of Neuroscience *27*, 419–451. https://doi.org/10.1146/annurev.neuro.27.070203.144152.

Driscoll, L.N., Pettit, N.L., Minderer, M., Chettih, S.N., and Harvey, C.D. (2017). Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. Cell *170*, 986-999.e16. https://doi.org/10.1016/j.cell.2017.07.021.

Driver, J., and Vuilleumier, P. (2001). Perceptual awareness and its loss in unilateral neglect and extinction. Cognition *79*, 39–88. https://doi.org/10.1016/S0010-0277(00)00124-4.

D'Souza, R.D., Bista, P., Meier, A.M., Ji, W., and Burkhalter, A. (2019). Spatial Clustering of Inhibition in Mouse Primary Visual Cortex. Neuron *104*, 588-600.e5. https://doi.org/10.1016/j.neuron.2019.09.020.

Durand, S., Iyer, R., Mizuseki, K., Vries, S. de, Mihalas, S., and Reid, R.C. (2016). A Comparison of Visual Response Properties in the Lateral Geniculate Nucleus and Primary Visual Cortex of Awake and Anesthetized Mice. J. Neurosci. *36*, 12144–12156. https://doi.org/10.1523/JNEUROSCI.1741-16.2016.

Ecker, A.S., Berens, P., Keliris, G.A., Bethge, M., Logothetis, N.K., and Tolias, A.S. (2010). Decorrelated Neuronal Firing in Cortical Microcircuits. Science *327*, 584–587. https://doi.org/10.1126/science.1179867.

Ecker, A.S., Berens, P., Cotton, R.J., Subramaniyan, M., Denfield, G.H., Cadwell, C.R., Smirnakis, S.M., Bethge, M., and Tolias, A.S. (2014). State Dependence of Noise Correlations in Macaque Primary Visual Cortex. Neuron *82*, 235–248. https://doi.org/10.1016/j.neuron.2014.02.006.

Engel, T.A., Chaisangmongkon, W., Freedman, D.J., and Wang, X.-J. (2015). Choice-correlated activity fluctuations underlie learning of neuronal category representation. Nature Communications *6*, 6454. https://doi.org/10.1038/ncomms7454.

Erisken, S., Vaiceliunaite, A., Jurjut, O., Fiorini, M., Katzner, S., and Busse, L. (2014). Effects of Locomotion Extend throughout the Mouse Early Visual System. Current Biology *24*, 2899–2907. https://doi.org/10.1016/j.cub.2014.10.045.

Erlich, J.C., Brunton, B.W., Duan, C.A., Hanks, T.D., and Brody, C.D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. ELife *4*, e05457. https://doi.org/10.7554/eLife.05457.

Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature *415*, 429–433. https://doi.org/10.1038/415429a.

Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical Evidence of Multimodal Integration in Primate Striate Cortex. J. Neurosci. *22*, 5749–5759. .

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47. .

Ferguson, K.A., and Cardin, J.A. (2020). Mechanisms underlying gain modulation in the cortex. Nat. Rev. Neurosci. *21*, 80–92. https://doi.org/10.1038/s41583-019-0253-y.

Fetsch, C.R., DeAngelis, G.C., and Angelaki, D.E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. Nat Rev Neurosci *14*, 429–442. https://doi.org/10.1038/nrn3503.

Fiser, A., Mahringer, D., Oyibo, H.K., Petersen, A.V., Leinweber, M., and Keller, G.B. (2016). Experience-dependent spatial expectations in mouse visual cortex. Nature Neuroscience *19*, 1658–1664. https://doi.org/10.1038/nn.4385.

Fishman, M.C., and Michael, C.R. (1973). Integration of auditory information in the cat's visual cortex. Vision Research *13*, 1415–1419. https://doi.org/10.1016/0042-6989(73)90002-3.

Flanigan, K.A.S., Wiegmann, D.D., Hebets, E.A., and Bingman, V.P. (2021). Multisensory integration supports configural learning of a home refuge in the whip spider Phrynus marginemaculatus. J Exp Biol *224*, jeb238444. https://doi.org/10.1242/jeb.238444.

Fong, M., Mitchell, D.E., Duffy, K.R., and Bear, M.F. (2016). Rapid recovery from the effects of early monocular deprivation is enabled by temporary inactivation of the retinas. PNAS *113*, 14139–14144. https://doi.org/10.1073/pnas.1613279113.

Freedman, D.J., and Assad, J.A. (2006). Experience-dependent representation of visual categories in parietal cortex. Nature *443*, 85–88. https://doi.org/10.1038/nature05078.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw *33*, 1–22. .

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends in Cognitive Sciences *9*, 474–480. https://doi.org/10.1016/j.tics.2005.08.011.

Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience *11*, 127–138. https://doi.org/10.1038/nrn2787.

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nature Neuroscience *6*, 1216–1223. https://doi.org/10.1038/nn1141.

Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., and Shamma, S.A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nature Neuroscience *13*, 1011–1019. https://doi.org/10.1038/nn.2598.

Froudarakis, E., Fahey, P.G., Reimer, J., Smirnakis, S.M., Tehovnik, E.J., and Tolias, A.S. (2019). The Visual Cortex in Context. Annu Rev Vis Sci *5*, 317–339. https://doi.org/10.1146/annurev-vision-091517-034407.

Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., and Stryker, M.P. (2014). A Cortical Circuit for Gain Control by Behavioral State. Cell *156*, 1139–1152. https://doi.org/10.1016/j.cell.2014.01.050.

Funamizu, A., Kuhn, B., and Doya, K. (2016). Neural substrate of dynamic Bayesian inference in the cerebral cortex. Nat Neurosci *19*, 1682–1689. https://doi.org/10.1038/nn.4390.

Gallero-Salas, Y., Han, S., Sych, Y., Voigt, F.F., Laurenczy, B., Gilad, A., and Helmchen, F. (2021). Sensory and Behavioral Components of Neocortical Signal Flow in Discrimination Tasks with Short-Term Memory. Neuron *109*, 135-148.e6. https://doi.org/10.1016/j.neuron.2020.10.017.

Gămănuţ, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D.C., Knoblauch, K., and Burkhalter, A. (2018). The Mouse Cortical Connectome, Characterized by an Ultra-Dense Cortical Graph, Maintains Specificity by Distinct Connectivity Profiles. Neuron *97*, 698-715.e10. https://doi.org/10.1016/j.neuron.2017.12.037.

Garner, A.R., and Keller, G.B. (2022). A cortical circuit for audio-visual predictions. Nat Neurosci *25*, 98–105. https://doi.org/10.1038/s41593-021-00974-7.

Gdalyahu, A., Tring, E., Polack, P.-O., Gruver, R., Golshani, P., Fanselow, M.S., Silva, A.J., and Trachtenberg, J.T. (2012). Associative Fear Learning Enhances Sparse Network Coding in Primary Sensory Cortex. Neuron *75*, 121–132. https://doi.org/10.1016/j.neuron.2012.04.035.

Gentet, L.J., Avermann, M., Matyas, F., Staiger, J.F., and Petersen, C.C.H. (2010). Membrane Potential Dynamics of GABAergic Neurons in the Barrel Cortex of Behaving Mice. Neuron *65*, 422–435. https://doi.org/10.1016/j.neuron.2010.01.006.

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

Ghose, G.M., Yang, T., and Maunsell, J.H.R. (2002). Physiological Correlates of Perceptual Learning in Monkey V1 and V2. Journal of Neurophysiology *87*, 1867–1888. https://doi.org/10.1152/jn.00690.2001.

Ghosh, D.D., Nitabach, M.N., Zhang, Y., and Harris, G. (2017). Multisensory integration in C. elegans. Current Opinion in Neurobiology *43*, 110–118. https://doi.org/10.1016/j.conb.2017.01.005.

Gilad, A., Gallero-Salas, Y., Groos, D., and Helmchen, F. (2018). Behavioral Strategy Determines Frontal or Posterior Location of Short-Term Memory in Neocortex. Neuron *99*, 814-828.e7. https://doi.org/10.1016/j.neuron.2018.07.029.

Gilbert, C.D., and Li, W. (2013). Top-down influences on visual processing. Nature Reviews Neuroscience *14*, 350–363. https://doi.org/10.1038/nrn3476.

Gleiss, S., and Kayser, C. (2012). Audio-Visual Detection Benefits in the Rat. PLoS One *7*. https://doi.org/10.1371/journal.pone.0045677.

Glickfeld, L.L., and Olsen, S.R. (2017). Higher-Order Areas of the Mouse Visual Cortex. Annu. Rev. Vis. Sci. *3*, 251–273. https://doi.org/10.1146/annurev-vision-102016-061331.

Glickfeld, L.L., Andermann, M.L., Bonin, V., and Reid, R.C. (2013a). Cortico-cortical projections in mouse visual cortex are functionally target specific. Nat Neurosci *16*, 219–226. https://doi.org/10.1038/nn.3300.

Glickfeld, L.L., Histed, M.H., and Maunsell, J.H.R. (2013b). Mouse Primary Visual Cortex Is Used to Detect Both Orientation and Contrast Changes. J. Neurosci. *33*, 19416–19422. https://doi.org/10.1523/JNEUROSCI.3560-13.2013.

Glickfeld, L.L., Reid, R.C., and Andermann, M.L. (2014). A mouse model of higher visual cortical function. Curr. Opin. Neurobiol. *24*, 28–33. https://doi.org/10.1016/j.conb.2013.08.009.

Goard, M., and Dan, Y. (2009). Basal forebrain activation enhances cortical coding of natural scenes. Nature Neuroscience *12*, 1444–1449. https://doi.org/10.1038/nn.2402.

Goard, M.J., Pho, G.N., Woodson, J., and Sur, M. (2016). Distinct roles of visual, parietal, and frontal motor cortices in memory-guided sensorimotor decisions. Elife *5*, e13764. .

Gold, J.I., and Shadlen, M.N. (2007). The Neural Basis of Decision Making. Annual Review of Neuroscience *30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038.

Golshani, P., Gonçalves, J.T., Khoshkhoo, S., Mostany, R., Smirnakis, S., and Portera-Cailliau, C. (2009). Internally Mediated Developmental Desynchronization of Neocortical Network Activity. J. Neurosci. *29*, 10890–10899. https://doi.org/10.1523/JNEUROSCI.2012-09.2009.

Goltstein, P.M., Coffey, E.B.J., Roelfsema, P.R., and Pennartz, C.M.A. (2013). In Vivo Two-Photon Ca2+ Imaging Reveals Selective Reward Effects on Stimulus-Specific Assemblies in Mouse Visual Cortex. J. Neurosci. *33*, 11540–11555. https://doi.org/10.1523/JNEUROSCI.1341-12.2013.

Goltstein, P.M., Montijn, J.S., and Pennartz, C.M.A. (2015). Effects of Isoflurane Anesthesia on Ensemble Patterns of Ca2+ Activity in Mouse V1: Reduced Direction Selectivity Independent of Increased Correlations in Cellular Activity. PLOS ONE *10*, e0118277. https://doi.org/10.1371/journal.pone.0118277.

Goltstein, P.M., Meijer, G.T., and Pennartz, C.M. (2018). Conditioning sharpens the spatial representation of rewarded stimuli in mouse primary visual cortex. ELife *7*, e37683. https://doi.org/10.7554/eLife.37683.

Green, D.M., and Swets, J.A. (1966). Signal detection theory and psychophysics (Oxford, England: John Wiley).

Greenberg, D.S., Houweling, A.R., and Kerr, J.N.D. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nature Neuroscience *11*, 749–751. https://doi.org/10.1038/nn.2140.

Gregory, R.L. (1980). Perceptions as hypotheses. Phil. Trans. R. Soc. Lond. B *290*, 181–197. https://doi.org/10.1098/rstb.1980.0090.

Grimes, J.A. (1996). On the Failure to Detect Changes in Scenes Across Saccades. In Perception, K. Akins, ed. (Oxford University Press), p.

Grunewald, A., Linden, J.F., and Andersen, R.A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area. I. Effects of training. J Neurophysiol *82*, 330–342. https://doi.org/10.1152/jn.1999.82.1.330.

Guitchounts, G., Masís, J., Wolff, S.B.E., and Cox, D. (2020). Encoding of 3D Head Orienting Movements in the Primary Visual Cortex. Neuron *108*, 512-525.e4. https://doi.org/10.1016/j.neuron.2020.07.014.

Guo, Z.V., Li, N., Huber, D., Ophir, E., Gutnisky, D., Ting, J.T., Feng, G., and Svoboda, K. (2014a). Flow of Cortical Activity Underlying a Tactile Decision in Mice. Neuron *81*, 179–194. https://doi.org/10.1016/j.neuron.2013.10.020.

Guo, Z.V., Hires, S.A., Li, N., O'Connor, D.H., Komiyama, T., Ophir, E., Huber, D., Bonardi, C., Morandell, K., Gutnisky, D., et al. (2014b). Procedures for Behavioral Experiments in Head-Fixed Mice. PLoS One *9*. https://doi.org/10.1371/journal.pone.0088678.

Haider, B., Duque, A., Hasenstaub, A.R., Yu, Y., and McCormick, D.A. (2007). Enhancement of Visual Responsiveness by Spontaneous Local Network Activity In Vivo. Journal of Neurophysiology *97*, 4186–4202. https://doi.org/10.1152/jn.01114.2006.

Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. Nature *493*, 97–100. https://doi.org/10.1038/nature11665.

Hajnal, M.A., Tran, D., Einstein, M., Martelo, M.V., Safaryan, K., Polack, P.-O., Golshani, P., and Orbán, G. (2021). Continuous multiplexed population representations of task context in the mouse primary visual cortex. BioRxiv https://doi.org/10.1101/2021.04.20.440666.

Halassa, M.M., and Kastner, S. (2017). Thalamic functions in distributed cognitive control. Nature Neuroscience *20*, 1669. https://doi.org/10.1038/s41593-017-0020-1.

Han, Y., Kebschull, J.M., Campbell, R.A.A., Cowan, D., Imhof, F., Zador, A.M., and Mrsic-Flogel, T.D. (2018). The logic of single-cell projections from visual cortex. Nature *556*, 51–56. https://doi.org/10.1038/nature26159.

Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., and Brody, C.D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature *520*, 220–223. https://doi.org/10.1038/nature14066.

Hansen, B.J., Chelaru, M.I., and Dragoi, V. (2012). Correlated Variability in Laminar Cortical Circuits. Neuron *76*, 590–602. https://doi.org/10.1016/j.neuron.2012.08.029.

Hanson, J.V.M., Whitaker, D., and Heron, J. (2009). Preferential processing of tactile events under conditions of divided attention. Neuroreport *20*, 1392–1396. https://doi.org/10.1097/WNR.0b013e3283319e25.

Harnad, S. (1990). The symbol grounding problem. Physica D: Nonlinear Phenomena *42*, 335–346. https://doi.org/10.1016/0167-2789(90)90087-6.

Harris, K.D., and Mrsic-Flogel, T.D. (2013). Cortical connectivity and sensory coding. Nature *503*, 51–58. https://doi.org/10.1038/nature12654.

Harris, K.D., and Shepherd, G.M.G. (2015). The neocortical circuit: themes and variations. Nat Neurosci *18*, 170–181. https://doi.org/10.1038/nn.3917.

Harris, J.A., Mihalas, S., Hirokawa, K.E., Whitesell, J.D., Choi, H., Bernard, A., Bohn, P., Caldejon, S., Casal, L., Cho, A., et al. (2019). Hierarchical organization of cortical and thalamic connectivity. Nature *575*, 195–202. https://doi.org/10.1038/s41586-019-1716-z.

Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature *484*, 62–68. https://doi.org/10.1038/nature10918.

Hasenstaub, A., Sachdev, R.N.S., and McCormick, D.A. (2007). State Changes Rapidly Modulate Cortical Neuronal Responsiveness. J. Neurosci. *27*, 9607–9622. https://doi.org/10.1523/JNEUROSCI.2184-07.2007.

Havenith, M.N., Zijderveld, P.M., van Heukelum, S., Abghari, S., Glennon, J.C., and Tiesinga, P. (2018). The Virtual-Environment-Foraging Task enables rapid training and single-trial metrics of attention in head-fixed mice. Sci Rep *8*, 17371. https://doi.org/10.1038/s41598-018-34966-8.

Heeger, D.J. (2017). Theory of cortical function. Proceedings of the National Academy of Sciences *114*, 1773–1782. https://doi.org/10.1073/pnas.1619788114.

Heesy, C.P. (2004). On the relationship between orbit orientation and binocular visual field overlap in mammals. The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology *281A*, 1104–1110. https://doi.org/10.1002/ar.a.20116.

Helmholtz, H. von (1867). Handbuch der physiologischen Optik (Leipzig : Leopold Voss).

Henry, K.R., and Lepkowski, C.M. (1978). Evoked Potential Correlates of Genetic Progressive Hearing Loss:Age-related Changes from the Ear to the Inferior Colliculus ofC57BL/6 and CBA/J Mice. Acta Oto-Laryngologica *86*, 366–374. https://doi.org/10.3109/00016487809124758.

Henschke, J.U., Noesselt, T., Scheich, H., and Budinger, E. (2015). Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. Brain Struct Funct *220*, 955–977. https://doi.org/10.1007/s00429-013-0694-4.

Henschke, J.U., Dylda, E., Katsanevaki, D., Dupuy, N., Currie, S.P., Amvrosiadis, T., Pakan, J.M.P., and Rochefort, N.L. (2020). Reward Association Enhances Stimulus-Specific Representations in Primary Visual Cortex. Current Biology *30*, 1866-1880.e5. https://doi.org/10.1016/j.cub.2020.03.018.

Hernández, A., Nácher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., Vázquez, Y., Camarillo, L., and Romo, R. (2010). Decoding a Perceptual Decision Process across Cortex. Neuron *66*, 300–314. https://doi.org/10.1016/j.neuron.2010.03.031.

Holmes, G. (1918). Disturbances of visual orientation. Br J Ophthalmol *2*, 449–468. .

Hong, Y.K., Lacefield, C.O., Rodgers, C.C., and Bruno, R.M. (2018). Sensation, movement and learning in the absence of barrel cortex. Nature *561*, 542–546. https://doi.org/10.1038/s41586-018-0527-y.

Hooks, B.M., and Chen, C. (2020). Circuitry Underlying Experience-Dependent Plasticity in the Mouse Visual System. Neuron *106*, 21–36. https://doi.org/10.1016/j.neuron.2020.01.031.

Hovde, K., Gianatti, M., Witter, M.P., and Whitlock, J.R. (2019). Architecture and organization of mouse posterior parietal cortex relative to extrastriate areas. European Journal of Neuroscience *49*, 1313–1329. https://doi.org/10.1111/ejn.14280.

Howard, I.P., and Templeton, W.B. (1966). Human spatial orientation.

Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. The Journal of Physiology *148*, 574–591. https://doi.org/10.1113/jphysiol.1959.sp006308.

Hubel, D.H., and Wiesel, T.N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol *28*, 229–289. https://doi.org/10.1152/jn.1965.28.2.229.

Hubel, D.H., and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. J. Physiol. (Lond.) *195*, 215–243. .

Hubel, D.H., and Wiesel, T.N. (1979). Brain Mechanisms of Vision. Scientific American *241*, 150–163. .

Hübener, M. (2003). Mouse visual cortex. Current Opinion in Neurobiology *13*, 413–420. https://doi.org/10.1016/S0959-4388(03)00102-8.

Huh, C.Y.L., Peach, J.P., Bennett, C., Vega, R.M., and Hestrin, S. (2018). Feature-Specific Organization of Feedback Pathways in Mouse Visual Cortex. Curr. Biol. *28*, 114-120.e5. https://doi.org/10.1016/j.cub.2017.11.056.

Hwang, E.J., Dahlen, J.E., Mukundan, M., and Komiyama, T. (2017). History-based action selection bias in posterior parietal cortex. Nature Communications *8*. https://doi.org/10.1038/s41467-017-01356-z.

Hyvärinen, J. (1982). Posterior parietal lobe of the primate brain. Physiological Reviews *62*, 1060–1129. https://doi.org/10.1152/physrev.1982.62.3.1060.

Ibrahim, Mesik, L., Ji, X., Fang, Q., Li, H., Li, Y., Zingg, B., Zhang, L.I., and Tao, H.W. (2016). Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition. Neuron *89*, 1031–1045. https://doi.org/10.1016/j.neuron.2016.01.027.

Issa, J.B., Haeffele, B.D., Agarwal, A., Bergles, D.E., Young, E.D., and Yue, D.T. (2014). Multiscale Optical Ca2+ Imaging of Tonal Organization in Mouse Auditory Cortex. Neuron *83*, 944–959. https://doi.org/10.1016/j.neuron.2014.07.009.

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., Tucci, V., Benfenati, F., and Medini, P. (2012). Sound-Driven Synaptic Inhibition in Primary Visual Cortex. Neuron *73*, 814–828. https://doi.org/10.1016/j.neuron.2011.12.026.

Jacobs, E.A.K., Steinmetz, N.A., Carandini, M., and Harris, K.D. (2018). Cortical state fluctuations during sensory decision making. BioRxiv 348193. https://doi.org/10.1101/348193.

Jazayeri, M., and Afraz, A. (2017). Navigating the Neural Space in Search of the Neural Code. Neuron *93*, 1003–1014. https://doi.org/10.1016/j.neuron.2017.02.019.

Jeanne, J.M., Sharpee, T.O., and Gentner, T.Q. (2013). Associative Learning Enhances Population Coding by Inverting Interneuronal Correlation Patterns. Neuron *78*, 352–363. https://doi.org/10.1016/j.neuron.2013.02.023.

Jeffreys, H. (1939). The Theory of Probability. The Theory of Probability, 3rd Edition, by Harold Jeffreys. Oxford Classic Texts in the Physical Sciences. ISBN: 9780198503682. Oxford: Oxford University Press, 1939.

Ji, W., Gămănuţ, R., Bista, P., D'Souza, R.D., Wang, Q., and Burkhalter, A. (2015). Modularity in the Organization of Mouse Primary Visual Cortex. Neuron *87*, 632–643. https://doi.org/10.1016/j.neuron.2015.07.004.

Jin, M., and Glickfeld, L.L. (2020). Mouse Higher Visual Areas Provide Both Distributed and Specialized Contributions to Visually Guided Behaviors. Current Biology *30*, 4682-4692.e7. https://doi.org/10.1016/j.cub.2020.09.015.

Jones, E.G., and Powell, T.P.S. (1970). An Anatomical Study of Converging Sensory Pathways Within the Cerebral Cortex of the Monkey. Brain *93*, 793–820. https://doi.org/10.1093/brain/93.4.793.

Jordan, R., and Keller, G.B. (2020). Opposing Influence of Top-down and Bottom-up Input on Excitatory Layer 2/3 Neurons in Mouse Primary Visual Cortex. Neuron *108*, 1194-1206.e5. https://doi.org/10.1016/j.neuron.2020.09.024.

Juavinett, A.L., Erlich, J.C., and Churchland, A.K. (2018). Decision-making behaviors: weighing ethology, complexity, and sensorimotor compatibility. Current Opinion in Neurobiology *49*, 42–50. https://doi.org/10.1016/j.conb.2017.11.001.

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*, nature24636. https://doi.org/10.1038/nature24636.

Kaas, J.H. (1980). A Comparative Survey of Visual Cortex Organization in Mammals. In Comparative Neurology of the Telencephalon, S.O.E. Ebbesson, ed. (Boston, MA: Springer US), pp. 483–502.

Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S., and Branson, K. (2013). JAABA: interactive machine learning for automatic annotation of animal behavior. Nat Methods *10*, 64–67. https://doi.org/10.1038/nmeth.2281.

Kaneko, M., Fu, Y., and Stryker, M.P. (2017). Locomotion Induces Stimulus-Specific Response Enhancement in Adult Visual Cortex. J. Neurosci. *37*, 3532–3543. https://doi.org/10.1523/JNEUROSCI.3760-16.2017.

Kaplan, H.S., and Zimmer, M. (2020). Brain-wide representations of ongoing behavior: a universal principle? Current Opinion in Neurobiology *64*, 60–69. https://doi.org/10.1016/j.conb.2020.02.008.

Katz, L.N., Yates, J.L., Pillow, J.W., and Huk, A.C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. Nature *535*, 285–288. https://doi.org/10.1038/nature18617.

Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2014). Cortical activity in the null space: permitting preparation without movement. Nature Neuroscience *17*, 440–448. https://doi.org/10.1038/nn.3643.

Kauvar, I.V., Machado, T.A., Yuen, E., Kochalka, J., Choi, M., Allen, W.E., Wetzstein, G., and Deisseroth, K. (2020). Cortical Observation by Synchronous Multifocal Optical Sampling Reveals Widespread Population Encoding of Actions. Neuron *107*, 351-367.e19. https://doi.org/10.1016/j.neuron.2020.04.023.

Kayser, C., and Logothetis, N.K. (2007). Do early sensory cortices integrate cross-modal information? Brain Struct Funct *212*, 121–132. https://doi.org/10.1007/s00429-007-0154-0.

Keller, G.B., and Mrsic-Flogel, T.D. (2018). Predictive Processing: A Canonical Cortical Computation. Neuron *100*, 424–435. https://doi.org/10.1016/j.neuron.2018.10.003.

Keller, A.J., Dipoppa, M., Roth, M.M., Caudill, M.S., Ingrosso, A., Miller, K.D., and Scanziani, M. (2020a). A Disinhibitory Circuit for Contextual Modulation in Primary Visual Cortex. Neuron *108*, 1181-1193.e8. https://doi.org/10.1016/j.neuron.2020.11.013.

Keller, A.J., Roth, M.M., and Scanziani, M. (2020b). Feedback generates a second receptive field in neurons of the visual cortex. Nature *582*, 545–549. https://doi.org/10.1038/s41586-020-2319-4.

Keller, G.B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse. Neuron *74*, 809–815. https://doi.org/10.1016/j.neuron.2012.03.040.

Kerkhoff, G. (2001). Spatial hemineglect in humans. Progress in Neurobiology *63*, 1–27. https://doi.org/10.1016/S0301-0082(00)00028-9.

Keysers, C., Gazzola, V., and Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. Nature Neuroscience *23*, 788–799. https://doi.org/10.1038/s41593-020-0660-4.

Khastkhodaei, Z., Jurjut, O., Katzner, S., and Busse, L. (2016). Mice Can Use Second-Order, Contrast-Modulated Stimuli to Guide Visual Perception. J. Neurosci. *36*, 4457–4469. https://doi.org/10.1523/JNEUROSCI.4595-15.2016.

Kim, M.-H., Znamenskiy, P., Iacaruso, M.F., and Mrsic-Flogel, T.D. (2018). Segregated Subnetworks of Intracortical Projection Neurons in Primary Visual Cortex. Neuron *100*, 1313-1321.e6. https://doi.org/10.1016/j.neuron.2018.10.023.

Kim, T.H., Zhang, Y., Lecoq, J., Jung, J.C., Li, J., Zeng, H., Niell, C.M., and Schnitzer, M.J. (2016). Long-Term Optical Access to an Estimated One Million Neurons in the Live Mouse Cortex. Cell Reports *17*, 3385–3394. https://doi.org/10.1016/j.celrep.2016.12.004.

Kirchberger, L., Mukherjee, S., Schnabel, U.H., Beest, E.H. van, Barsegyan, A., Levelt, C.N., Heimel, J.A., Lorteije, J.A.M., Togt, C. van der, Self, M.W., et al. (2021). The essential role of recurrent processing for figure-ground perception in mice. Science Advances *7*, eabe1833. https://doi.org/10.1126/sciadv.abe1833.

Knill, D.C., and Richards, W. (1996). Perception as Bayesian Inference (Cambridge University Press).

Knöpfel, T., Sweeney, Y., Radulescu, C.I., Zabouri, N., Doostdar, N., Clopath, C., and Barnes, S.J. (2019). Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex. Nature Communications *10*, 5684. https://doi.org/10.1038/s41467-019-13607-2.

Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjöström, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. Nature *473*, 87–91. https://doi.org/10.1038/nature09880.

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. Nat. Rev. Neurosci. *17*, 307–321. https://doi.org/10.1038/nrn.2016.22.

Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and Neuronal Population Information. Annual Review of Neuroscience *39*, 237–256. https://doi.org/10.1146/annurev-neuro-070815-013851.

Kohn, A., Jasper, A.I., Semedo, J.D., Gokcen, E., Machens, C.K., and Yu, B.M. (2020). Principles of Corticocortical Communication: Proposed Schemes and Design Considerations. Trends in Neurosciences *43*, 725–737. https://doi.org/10.1016/j.tins.2020.07.001.

Kok, P., Failing, M.F., and de Lange, F.P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. Journal of Cognitive Neuroscience *26*, 1546–1554. https://doi.org/10.1162/jocn_a_00562.

Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B., and Shams, L. (2007). Causal Inference in Multisensory Perception. PLOS ONE *2*, e943. https://doi.org/10.1371/journal.pone.0000943.

Korthauer, K., Kimes, P.K., Duvallet, C., Reyes, A., Subramanian, A., Teng, M., Shukla, C., Alm, E.J., and Hicks, S.C. (2019). A practical guide to methods controlling false discoveries in computational biology. Genome Biol *20*, 118. https://doi.org/10.1186/s13059-019-1716-1.

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron *93*, 480–490. https://doi.org/10.1016/j.neuron.2016.12.041.

Krauzlis, R.J., Lovejoy, L.P., and Zénon, A. (2013). Superior Colliculus and Visual Spatial Attention. Annual Review of Neuroscience *36*, 165–182. https://doi.org/10.1146/annurev-neuro-062012-170249.

Krauzlis, R.J., Bogadhi, A.R., Herman, J.P., and Bollimunta, A. (2018). Selective attention without a neocortex. Cortex *102*, 161–175. https://doi.org/10.1016/j.cortex.2017.08.026.

Krumin, M., Lee, J.J., Harris, K.D., and Carandini, M. (2018). Decision and navigation in mouse parietal cortex. ELife *7*, e42583. https://doi.org/10.7554/eLife.42583.

Lacefield, C.O., Pnevmatikakis, E.A., Paninski, L., and Bruno, R.M. (2019). Reinforcement Learning Recruits Somata and Apical Dendrites across Layers of Primary Sensory Cortex. Cell Reports *26*, 2000-2008.e2. https://doi.org/10.1016/j.celrep.2019.01.093.

Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C., and Schroeder, C.E. (2009). The Leading Sense: Supramodal Control of Neurophysiological Context by Attention. Neuron *64*, 419–430. https://doi.org/10.1016/j.neuron.2009.10.014.

Lamme, V.A.F., and Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends in Neurosciences *23*, 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X.

Lamme, V.A., Zipser, K., and Spekreijse, H. (1998). Figure-ground activity in primary visual cortex is suppressed by anesthesia. Proc Natl Acad Sci U S A *95*, 3263–3268. https://doi.org/10.1073/pnas.95.6.3263.

Lampl, I., Reichova, I., and Ferster, D. (1999). Synchronous Membrane Potential Fluctuations in Neurons of the Cat Visual Cortex. Neuron *22*, 361–374. https://doi.org/10.1016/S0896-6273(00)81096-X.

Lauer, S.M., Schneeweiß, U., Brecht, M., and Ray, S. (2018). Visualization of Cortical Modules in Flattened Mammalian Cortices. JoVE (Journal of Visualized Experiments) e56992. https://doi.org/10.3791/56992.

Le Merre, P., Esmaeili, V., Charrière, E., Galan, K., Salin, P.-A., Petersen, C.C.H., and Crochet, S. (2018). Reward-Based Learning Drives Rapid Sensory Signals in Medial Prefrontal Cortex and Dorsal Hippocampus Necessary for Goal-Directed Behavior. Neuron *97*, 83-91.e5. https://doi.org/10.1016/j.neuron.2017.11.031.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444. https://doi.org/10.1038/nature14539.

Lee, T.S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. J. Opt. Soc. Am. A, JOSAA *20*, 1434–1448. https://doi.org/10.1364/JOSAA.20.001434.

Lee, C.C.Y., Diamond, M.E., and Arabzadeh, E. (2016). Sensory Prioritization in Rats: Behavioral Performance and Neuronal Correlates. J. Neurosci. *36*, 3243–3253. https://doi.org/10.1523/JNEUROSCI.3636-15.2016.

Lee, S., Hjerling-Leffler, J., Zagha, E., Fishell, G., and Rudy, B. (2010). The Largest Group of Superficial Neocortical GABAergic Interneurons Expresses Ionotropic Serotonin Receptors. J. Neurosci. *30*, 16796–16808. https://doi.org/10.1523/JNEUROSCI.1869-10.2010.

Lehman, E.B. (1982). Memory for modality: Evidence for an automatic process. Mem Cogn *10*, 554–564. https://doi.org/10.3758/BF03202438.

Leinweber, M., Ward, D.R., Sobczak, J.M., Attinger, A., and Keller, G.B. (2017). A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions. Neuron *95*, 1420-1432.e5. https://doi.org/10.1016/j.neuron.2017.08.036.

Levelt, C.N., and Hübener, M. (2012). Critical-Period Plasticity in the Visual Cortex. Annual Review of Neuroscience *35*, 309–330. https://doi.org/10.1146/annurev-neuro-061010-113813.

Lewicki, M.S. (2002). Efficient coding of natural sounds. Nat Neurosci *5*, 356–363. https://doi.org/10.1038/nn831.

Li, N., Chen, T.-W., Guo, Z.V., Gerfen, C.R., and Svoboda, K. (2015). A motor cortex circuit for motor planning and movement. Nature *519*, 51–56. https://doi.org/10.1038/nature14178.

Li, N., Chen, S., Guo, Z.V., Chen, H., Huo, Y., Inagaki, H.K., Chen, G., Davis, C., Hansel, D., Guo, C., et al. (2019). Spatiotemporal constraints on optogenetic inactivation in cortical circuits. ELife *8*, e48622. https://doi.org/10.7554/eLife.48622.

Li, Y., Ibrahim, L.A., Liu, B., Zhang, L.I., and Tao, H.W. (2013). Linear transformation of thalamocortical input by intracortical excitation. Nat Neurosci *16*, 1324–1330. https://doi.org/10.1038/nn.3494.

Liang, H., Gong, X., Chen, M., Yan, Y., Li, W., and Gilbert, C.D. (2017). Interactions between feedback and lateral connections in the primary visual cortex. PNAS 201706183. https://doi.org/10.1073/pnas.1706183114.

Licata, A.M., Kaufman, M.T., Raposo, D., Ryan, M.B., Sheppard, J.P., and Churchland, A.K. (2017). Posterior Parietal Cortex Guides Visual Decisions in Rats. J. Neurosci. *37*, 4954–4966. https://doi.org/10.1523/JNEUROSCI.0105-17.2017.

Lien, A.D., and Scanziani, M. (2013). Tuned thalamic excitation is amplified by visual cortical circuits. Nat Neurosci *16*, 1315–1323. https://doi.org/10.1038/nn.3488.

Linderman, S.W., and Gershman, S.J. (2017). Using computational theory to constrain statistical models of neural data. Current Opinion in Neurobiology *46*, 14–24. https://doi.org/10.1016/j.conb.2017.06.004.

Lippert, M., Logothetis, N.K., and Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. Brain Research *1173*, 102–109. https://doi.org/10.1016/j.brainres.2007.07.050.

Livingstone, M.S., and Hubel, D.H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. J. Neurosci. *7*, 3416–3468. https://doi.org/10.1523/JNEUROSCI.07-11-03416.1987.

Logothetis, N.K., Kayser, C., and Oeltermann, A. (2007). In Vivo Measurement of Cortical Impedance Spectrum in Monkeys: Implications for Signal Propagation. Neuron *55*, 809–823. https://doi.org/10.1016/j.neuron.2007.07.027.

Lyamzin, D., and Benucci, A. (2019). The mouse posterior parietal cortex: Anatomy and functions. Neuroscience Research *140*, 14–22. https://doi.org/10.1016/j.neures.2018.10.008.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nature Neuroscience *13*, 133–140. https://doi.org/10.1038/nn.2467.

Madisen, L., Mao, T., Koch, H., Zhuo, J., Berenyi, A., Fujisawa, S., Hsu, Y.-W.A., Garcia, A.J., 3rd, Gu, X., Zanella, S., et al. (2012). A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. Nat. Neurosci. *15*, 793–802. https://doi.org/10.1038/nn.3078.

Maheswaranathan, N., Kastner, D.B., Baccus, S.A., and Ganguli, S. (2018). Inferring hidden structure in multilayered neural circuits. PLOS Computational Biology *14*, e1006291. https://doi.org/10.1371/journal.pcbi.1006291.

Maier, J.X., Blankenship, M.L., Li, J.X., and Katz, D.B. (2015). A Multisensory Network for Olfactory Processing. Current Biology *25*, 2642–2650. https://doi.org/10.1016/j.cub.2015.08.060.

Makino, H., and Komiyama, T. (2015). Learning enhances the relative impact of top-down processing in the visual cortex. Nat. Neurosci. *18*, 1116–1122. https://doi.org/10.1038/nn.4061.

Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., Ota, K., Matsubara, C., Inutsuka, A., Sato, M., et al. (2015). A Top-Down Cortical Circuit for Accurate Sensory Perception. Neuron *86*, 1304–1316. https://doi.org/10.1016/j.neuron.2015.05.006.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84. https://doi.org/10.1038/nature12742.

Mao, T., Kusefoglu, D., Hooks, B.M., Huber, D., Petreanu, L., and Svoboda, K. (2011). Long-Range Neuronal Circuits Underlying the Interaction between Sensory and Motor Cortex. Neuron *72*, 111–123. https://doi.org/10.1016/j.neuron.2011.07.029.

Mariño, J., Schummers, J., Lyon, D.C., Schwabe, L., Beck, O., Wiesing, P., Obermayer, K., and Sur, M. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. Nat Neurosci *8*, 194–201. https://doi.org/10.1038/nn1391.

Markov, N.T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., et al. (2014). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. J. Comp. Neurol. *522*, 225–259. https://doi.org/10.1002/cne.23458.

Marques, T., Nguyen, J., Fioreze, G., and Petreanu, L. (2018). The functional organization of cortical feedback inputs to primary visual cortex. Nat. Neurosci. *21*, 757–764. https://doi.org/10.1038/s41593-018-0135-z.

Marr, D., and Hildreth, E. (1980). Theory of edge detection. Proceedings of the Royal Society of London. Series B. Biological Sciences *207*, 187–217. https://doi.org/10.1098/rspb.1980.0020.

Marshel, J.H., Garrett, M.E., Nauhaus, I., and Callaway, E.M. (2011). Functional specialization of seven mouse visual cortical areas. Neuron *72*, 1040–1054. https://doi.org/10.1016/j.neuron.2011.12.004.

Martínez, A., Anllo-Vento, L., Sereno, M.I., Frank, L.R., Buxton, R.B., Dubowitz, D.J., Wong, E.C., Hinrichs, H., Heinze, H.J., and Hillyard, S.A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat Neurosci *2*, 364–369. https://doi.org/10.1038/7274.

Masset, P., Ott, T., Lak, A., Hirokawa, J., and Kepecs, A. (2020). Behavior- and Modality-General Representation of Confidence in Orbitofrontal Cortex. Cell *182*, 112-126.e18. https://doi.org/10.1016/j.cell.2020.05.022.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature Neuroscience *21*, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y.

Mathot, S. (2018). Pupillometry: Psychology, Physiology, and Function. Journal of Cognition *1*, 16. https://doi.org/10.5334/joc.18.

Matteucci, G., Bellacosa Marotti, R., Riggi, M., Rosselli, F.B., and Zoccolan, D. (2019). Nonlinear Processing of Shape Information in Rat Lateral Extrastriate Cortex. J. Neurosci. *39*, 1649–1670. https://doi.org/10.1523/JNEUROSCI.1938-18.2018.

Maunsell, J.H.R. (2004). Neuronal representations of cognitive state: reward or attention? Trends in Cognitive Sciences *8*, 261–265. https://doi.org/10.1016/j.tics.2004.04.003.

Maunsell, J.H.R. (2015). Neuronal Mechanisms of Visual Attention. Annual Review of Vision Science *1*, 373–391. https://doi.org/10.1146/annurev-vision-082114-035431.

Mayner, W.G.P., Marshall, W., Billeh, Y.N., Gandhi, S.R., Caldejon, S., Cho, A., Griffin, F., Hancock, N., Lambert, S., Lee, E.K., et al. (2022). Measuring Stimulus-Evoked Neurophysiological Differentiation in Distinct Populations of Neurons in Mouse Visual Cortex. ENeuro *9*. https://doi.org/10.1523/ENEURO.0280-21.2021.

Mayrhofer, J.M., El-Boustani, S., Foustoukos, G., Auffret, M., Tamura, K., and Petersen, C.C.H. (2019). Distinct Contributions of Whisker Sensory Cortex and Tongue-Jaw Motor Cortex in a Goal-Directed Sensorimotor Transformation. Neuron *103*, 1034-1043.e5. https://doi.org/10.1016/j.neuron.2019.07.008.

McAdams, C.J., and Reid, R.C. (2005). Attention Modulates the Responses of Simple Cells in Monkey Primary Visual Cortex. J. Neurosci. *25*, 11023–11033. https://doi.org/10.1523/JNEUROSCI.2904-05.2005.

McCormick, D.A., Nestvogel, D.B., and He, B.J. (2020). Neuromodulation of Brain State and Behavior. Annual Review of Neuroscience *43*, 391–415. https://doi.org/10.1146/annurev-neuro-100219-105424.

McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., and McCormick, D.A. (2015a). Waking State: Rapid Variations Modulate Neural and Behavioral Responses. Neuron *87*, 1143–1161. https://doi.org/10.1016/j.neuron.2015.09.012.

McGinley, M.J., David, S.V., and McCormick, D.A. (2015b). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. Neuron *87*, 179–192. https://doi.org/10.1016/j.neuron.2015.05.038.

Mcgurk, H., and Macdonald, J. (1976). Hearing lips and seeing voices. Nature *264*, 746–748. https://doi.org/10.1038/264746a0.

Mehrotra, R., Namuduri, K.R., and Ranganathan, N. (1992). Gabor filter-based edge detection. Pattern Recognition *25*, 1479–1494. https://doi.org/10.1016/0031-3203(92)90121-X.

Mehta, A.D., Ulbert, I., and Schroeder, C.E. (2000). Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. Cereb Cortex *10*, 343–358. https://doi.org/10.1093/cercor/10.4.343.

Meijer, G.T., Montijn, J.S., Pennartz, C.M.A., and Lansink, C.S. (2017). Audio-visual modulation in mouse V1 depends on cross-modal stimulus configuration and congruency. J. Neurosci. https://doi.org/10.1523/JNEUROSCI.0468-17.2017.

Meijer, G.T., Pie, J.L., Dolman, T.L., Pennartz, C.M.A., and Lansink, C.S. (2018). Audiovisual Integration Enhances Stimulus Detection Performance in Mice. Front. Behav. Neurosci. *12*. https://doi.org/10.3389/fnbeh.2018.00231.

Meijer, G.T., Mertens, P.E.C., Pennartz, C.M.A., Olcese, U., and Lansink, C.S. (2019). The circuit architecture of cortical multisensory processing: Distinct functions jointly operating within a common anatomical network. Progress in Neurobiology *174*, 1–15. https://doi.org/10.1016/j.pneurobio.2019.01.004.

Meijer, G.T., Marchesi, P., Mejias, J.F., Montijn, J.S., Lansink, C.S., and Pennartz, C.M.A. (2020). Neural Correlates of Multisensory Detection Behavior: Comparison of Primary and Higher-Order Visual Cortex. Cell Reports *31*, 107636. https://doi.org/10.1016/j.celrep.2020.107636.

Mejías, J.F., and Wang, X.-J. (2022). Mechanisms of distributed working memory in a large-scale network of macaque neocortex. ELife *11*, e72136. https://doi.org/10.7554/eLife.72136.

Meredith, M.A., and Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. Journal of Neurophysiology *56*, 640–662. .

Mesik, L., Huang, J.J., Zhang, L.I., and Tao, H.W. (2019). Sensory- and Motor-Related Responses of Layer 1 Neurons in the Mouse Visual Cortex. J. Neurosci. *39*, 10060–10070. https://doi.org/10.1523/JNEUROSCI.1722-19.2019.

Mihalik, A., and Noppeney, U. (2020). Causal Inference in Audiovisual Perception. J. Neurosci. *40*, 6600–6612. https://doi.org/10.1523/JNEUROSCI.0051-20.2020.

Miller, E.K., and Cohen, J.D. (2001). An Integrative Theory of Prefrontal Cortex Function. Annual Review of Neuroscience *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167.

Miller, M.W., and Vogt, B.A. (1984). Direct connections of rat visual cortex with sensory, motor, and association cortices. Journal of Comparative Neurology *226*, 184–202. .

Mimica, B., Dunn, B.A., Tombaz, T., Bojja, V.P.T.N.C.S., and Whitlock, J.R. (2018). Efficient cortical coding of 3D posture in freely behaving rats. Science *362*, 584–589. https://doi.org/10.1126/science.aau2013.

Minamimoto, T., and Kimura, M. (2002). Participation of the Thalamic CM-Pf Complex in Attentional Orienting. Journal of Neurophysiology *87*, 3090–3101. https://doi.org/10.1152/jn.2002.87.6.3090.

Mishkin, M., Ungerleider, L.G., and Macko, K.A. (1983). Object vision and spatial vision: two cortical pathways. Trends in Neurosciences *6*, 414–417. https://doi.org/10.1016/0166-2236(83)90190-X.

Mochol, G., Hermoso-Mendizabal, A., Sakata, S., Harris, K.D., and Rocha, J. de la (2015). Stochastic transitions into silence cause noise correlations in cortical circuits. PNAS *112*, 3529–3534. https://doi.org/10.1073/pnas.1410509112.

Mohajerani, M.H., Aminoltejari, K., and Murphy, T.H. (2011). Targeted mini-strokes produce changes in interhemispheric sensory signal processing that are indicative of disinhibition within minutes. Proc. Natl. Acad. Sci. U.S.A. *108*, E183-191. https://doi.org/10.1073/pnas.1101914108.

Mohan, H., Gallero-Salas, Y., Carta, S., Sacramento, J., Laurenczy, B., Sumanovski, L.T., Kock, C.P.J. de, Helmchen, F., and Sachidhanandam, S. (2018). Sensory representation of an auditory cued tactile stimulus in the posterior parietal cortex of the mouse. Sci Rep *8*, 1–13. https://doi.org/10.1038/s41598-018-25891-x.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., and Foxe, J.J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cognitive Brain Research *14*, 115–128. https://doi.org/10.1016/S0926-6410(02)00066-6.

Montijn, J.S., Vinck, M., and Pennartz, C.M.A. (2014). Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation. Front. Comput. Neurosci. *8*. https://doi.org/10.3389/fncom.2014.00058.

Montijn, J.S., Goltstein, P.M., and Pennartz, C.M. (2015). Mouse V1 population correlates of visual detection rely on heterogeneity within neuronal response patterns. ELife *4*, e10163. https://doi.org/10.7554/eLife.10163.

Montijn, J.S., Meijer, G.T., Lansink, C.S., and Pennartz, C.M.A. (2016). Population-Level Neural Codes Are Robust to Single-Neuron Variability from a Multidimensional Coding Perspective. Cell Reports *16*, 2486–2498. https://doi.org/10.1016/j.celrep.2016.07.065.

Montijn, J.S., Seignette, K., Howlett, M.H., Cazemier, J.L., Kamermans, M., Levelt, C.N., and Heimel, J.A. (2021). A parameter-free statistical test for neuronal responsiveness. ELife *10*, e71969. https://doi.org/10.7554/eLife.71969.

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nature Neuroscience *17*, 1410–1417. https://doi.org/10.1038/nn.3807.

Morrell, F. (1972). Visual System's View of Acoustic Space. Nature *238*, 44–46. https://doi.org/10.1038/238044a0.

Mumford, D. (1992). On the computational architecture of the neocortex. Biol. Cybern. *66*, 241–251. https://doi.org/10.1007/BF00198477.

Murgas, K.A., Wilson, A.M., Michael, V., and Glickfeld, L.L. (2020). Unique Spatial Integration in Mouse Primary Visual Cortex and Higher Visual Areas. J. Neurosci. *40*, 1862–1873. https://doi.org/10.1523/JNEUROSCI.1997-19.2020.

Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nature Neuroscience *22*, 1677–1686. https://doi.org/10.1038/s41593-019-0502-4.

Najafi, F., Elsayed, G.F., Cao, R., Pnevmatikakis, E., Latham, P.E., Cunningham, J.P., and Churchland, A.K. (2020). Excitatory and Inhibitory Subnetworks Are Equally Selective during Decision-Making and Emerge Simultaneously during Learning. Neuron *105*, 165-179.e8. https://doi.org/10.1016/j.neuron.2019.09.045.

Nandy, A.S., Nassi, J.J., and Reynolds, J.H. (2016). Laminar Organization of Attentional Modulation in Macaque Visual Area V4. Neuron https://doi.org/10.1016/j.neuron.2016.11.029.

Nassi, J.J., Lomber, S.G., and Born, R.T. (2013). Corticocortical Feedback Contributes to Surround Suppression in V1 of the Alert Primate. J. Neurosci. *33*, 8504–8517. https://doi.org/10.1523/JNEUROSCI.5124-12.2013.

Ni, A.M., Ruff, D.A., Alberts, J.J., Symmonds, J., and Cohen, M.R. (2018). Learning and attention reveal a general relationship between population activity and behavior. Science *359*, 463–465. https://doi.org/10.1126/science.aao0284.

Nieder, A., Wagener, L., and Rinnert, P. (2020). A neural correlate of sensory consciousness in a corvid bird. Science *369*, 1626–1629. https://doi.org/10.1126/science.abb1447.

Niell, C.M., and Stryker, M.P. (2008). Highly Selective Receptive Fields in Mouse Visual Cortex. J Neurosci *28*, 7520–7536. https://doi.org/10.1523/JNEUROSCI.0623-08.2008.

Niell, C.M., and Stryker, M.P. (2010). Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. Neuron *65*, 472–479. https://doi.org/10.1016/j.neuron.2010.01.033.

Nikbakht, N., Tafreshiha, A., Zoccolan, D., and Diamond, M.E. (2018). Supralinear and Supramodal Integration of Visual and Tactile Signals in Rats: Psychophysics and Neuronal Mechanisms. Neuron *97*, 626–639. https://doi.org/10.1016/j.neuron.2018.01.003.

Nir, Y., Vyazovskiy, V.V., Cirelli, C., Banks, M.I., and Tononi, G. (2015). Auditory Responses and Stimulus-Specific Adaptation in Rat Auditory Cortex are Preserved Across NREM and REM Sleep. Cereb Cortex *25*, 1362–1378. https://doi.org/10.1093/cercor/bht328.

Noesselt, T., Hillyard, S.A., Woldorff, M.G., Schoenfeld, A., Hagner, T., Jäncke, L., Tempelmann, C., Hinrichs, H., and Heinze, H.-J. (2002). Delayed striate cortical activation during spatial attention. Neuron *35*, 575–587. https://doi.org/10.1016/s0896-6273(02)00781-x.

Noppeney, U. (2021). Perceptual Inference, Learning, and Attention in a Multisensory World. Annual Review of Neuroscience *44*, 449–473. https://doi.org/10.1146/annurev-neuro-100120-085519.

Nurminen, L., Merlin, S., Bijanzadeh, M., Federer, F., and Angelucci, A. (2018). Top-down feedback controls spatial summation and response amplitude in primate visual cortex. Nature Communications *9*, 1–13. https://doi.org/10.1038/s41467-018-04500-5.

Odoemene, O., Pisupati, S., Nguyen, H., and Churchland, A.K. (2018). Visual Evidence Accumulation Guides Decision-Making in Unrestrained Mice. J. Neurosci. *38*, 10143–10155. https://doi.org/10.1523/JNEUROSCI.3478-17.2018.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). A mesoscale connectome of the mouse brain. Nature *508*, 207–214. https://doi.org/10.1038/nature13186.

Olcese, U., Iurilli, G., and Medini, P. (2013). Cellular and Synaptic Architecture of Multisensory Integration in the Mouse Neocortex. Neuron *79*, 579–593. https://doi.org/10.1016/j.neuron.2013.06.010.

Olcese, U., Bos, J.J., Vinck, M., Lankelma, J.V., van Mourik-Donga, L.B., Schlumm, F., and Pennartz, C.M.A. (2016). Spike-Based Functional Connectivity in Cerebral Cortex and Hippocampus: Loss of Global Connectivity Is Coupled to Preservation of Local Connectivity During Non-REM Sleep. J. Neurosci. *36*, 7676–7692. https://doi.org/10.1523/JNEUROSCI.4201-15.2016.

Olcese, U., Oude Lohuis, M., and Pennartz, C. (2018). Sensory processing across conscious and nonconscious brain states: from single neurons to distributed networks for inferential representation. Front. Syst. Neurosci. *12*. https://doi.org/10.3389/fnsys.2018.00049.

Olsen, S.R., Bortone, D.S., Adesnik, H., and Scanziani, M. (2012). Gain control by layer six in cortical circuits of vision. Nature *483*, 47–52. https://doi.org/10.1038/nature10835.

Olshausen, B.A., and Field, D.J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Research *37*, 3311–3325. https://doi.org/10.1016/S0042-6989(97)00169-7.

Otchy, T.M., Wolff, S.B.E., Rhee, J.Y., Pehlevan, C., Kawai, R., Kempf, A., Gobes, S.M.H., and Ölveczky, B.P. (2015). Acute off-target effects of neural circuit manipulations. Nature *528*, 358–363. https://doi.org/10.1038/nature16442.

Oude Lohuis, M.N., Pie, J.L., Marchesi, P., Montijn, J.S., de Kock, C.P.J., Pennartz, C.M.A., and Olcese, U. (2022). Multisensory task demands temporally extend the causal requirement for visual cortex in perception. Nat Commun *13*, 2864. https://doi.org/10.1038/s41467-022-30600-4.

Overgaard, M., Lindeløv, J., Svejstrup, S., Døssing, M., Hvid, T., Kauffmann, O., and Mouridsen, K. (2013). Is Conscious Stimulus Identification Dependent on Knowledge of the Perceptual Modality? Testing the "Source Misidentification Hypothesis." Front. Psychol. *4*. https://doi.org/10.3389/fpsyg.2013.00116.

Pachitariu, M., Lyamzin, D.R., Sahani, M., and Lesica, N.A. (2015). State-Dependent Population Coding in Primary Auditory Cortex. Journal of Neuroscience *35*, 2058–2073. https://doi.org/10.1523/JNEUROSCI.3318-14.2015.

Pafundo, D.E., Nicholas, M.A., Zhang, R., and Kuhlman, S.J. (2016). Top-Down-Mediated Facilitation in the Visual Cortex Is Gated by Subcortical Neuromodulation. J. Neurosci. *36*, 2904–2914. https://doi.org/10.1523/JNEUROSCI.2909-15.2016.

Pak, A., Ryu, E., Li, C., and Chubykin, A.A. (2020). Top-Down Feedback Controls the Cortical Representation of Illusory Contours in Mouse Primary Visual Cortex. J. Neurosci. *40*, 648–660. https://doi.org/10.1523/JNEUROSCI.1998-19.2019.

Panzeri, S., Harvey, C.D., Piasini, E., Latham, P.E., and Fellin, T. (2017). Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior. Neuron *93*, 491–507. https://doi.org/10.1016/j.neuron.2016.12.036.

Paperna, T., and Malach, R. (1991). Patterns of sensory intermodality relationships in the cerebral cortex of the rat. Journal of Comparative Neurology *308*, 432–456. .

Park, I.M., Meister, M.L.R., Huk, A.C., and Pillow, J.W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. Nature Neuroscience *17*, 1395–1403. https://doi.org/10.1038/nn.3800.

Parker, P.R.L., Brown, M.A., Smear, M.C., and Niell, C.M. (2020). Movement-Related Signals in Sensory Areas: Roles in Natural Behavior. Trends in Neurosciences *43*, 581–595. https://doi.org/10.1016/j.tins.2020.05.005.

Paxinos, G., and Franklin, K.B.J. (2004). The Mouse Brain in Stereotaxic Coordinates (Gulf Professional Publishing).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Louppe, G., Prettenhofer, P., Weiss, R., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830. .

Pennartz, C.M. (2015). The brain's representational power: on consciousness and the integration of modalities (MIT Press).

Pennartz, C.M.A. (2009). Identification and integration of sensory modalities: Neural basis and relation to consciousness. Consciousness and Cognition *18*, 718–739. https://doi.org/10.1016/j.concog.2009.03.003.

Pennartz, C.M.A., Dora, S., Muckli, L., and Lorteije, J.A.M. (2019). Towards a Unified View on Pathways and Functions of Neural Recurrent Processing. Trends in Neurosciences *42*, 589–603. https://doi.org/10.1016/j.tins.2019.07.005.

Petersen, C.C.H., Hahn, T.T.G., Mehta, M., Grinvald, A., and Sakmann, B. (2003). Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. Proc. Natl. Acad. Sci. U.S.A. *100*, 13638–13643. https://doi.org/10.1073/pnas.2235811100.

Petreanu, L., Mao, T., Sternson, S.M., and Svoboda, K. (2009). The subcellular organization of neocortical excitatory connections. Nature *457*, 1142–1145. https://doi.org/10.1038/nature07709.

Petro, L.S., Paton, A.T., and Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. Philosophical Transactions of the Royal Society B: Biological Sciences *372*, 20160104. https://doi.org/10.1098/rstb.2016.0104.

Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., and Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. Nature Neuroscience *16*, 1068–1076. https://doi.org/10.1038/nn.3446.

Philiastides, M.G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. J Neurosci *26*, 8965–8975. https://doi.org/10.1523/JNEUROSCI.1655-06.2006.

Pho, G.N., Goard, M.J., Woodson, J., Crawford, B., and Sur, M. (2018). Task-dependent representations of stimulus and choice in mouse parietal cortex. Nat Commun *9*, 1–16. https://doi.org/10.1038/s41467-018-05012-y.

Pinto, L., Goard, M.J., Estandian, D., Xu, M., Kwan, A.C., Lee, S.-H., Harrison, T.C., Feng, G., and Dan, Y. (2013). Fast modulation of visual perception by basal forebrain cholinergic neurons. Nature Neuroscience *16*, 1857–1863. https://doi.org/10.1038/nn.3552.

Pinto, L., Rajan, K., DePasquale, B., Thiberge, S.Y., Tank, D.W., and Brody, C.D. (2019). Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions. Neuron *104*, 810-824.e9. https://doi.org/10.1016/j.neuron.2019.08.025.

Platt, M.L., and Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex. Nature *400*, 233–238. https://doi.org/10.1038/22268.

Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state–dependent gain modulation in visual cortex. Nature Neuroscience *16*, 1331–1339. https://doi.org/10.1038/nn.3464.

Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G.B., Mrsic-Flogel, T.D., et al. (2015). Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. Neuron *86*, 1478–1490. https://doi.org/10.1016/j.neuron.2015.05.037.

Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. Nature Reviews Neuroscience *1*, 125–132. https://doi.org/10.1038/35039062.

Poulet, J.F.A., and Petersen, C.C.H. (2008). Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. Nature *454*, 881–885. https://doi.org/10.1038/nature07150.

Rao, R.P.N., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci *2*, 79–87. https://doi.org/10.1038/4580.

Raposo, D., Kaufman, M.T., and Churchland, A.K. (2014). A category-free neural population supports evolving demands during decision-making. Nat Neurosci *17*, 1784–1792. https://doi.org/10.1038/nn.3865.

Reimer, J., Froudarakis, E., Cadwell, C.R., Yatsenko, D., Denfield, G.H., and Tolias, A.S. (2014). Pupil Fluctuations Track Fast Switching of Cortical States during Quiet Wakefulness. Neuron *84*, 355–362. https://doi.org/10.1016/j.neuron.2014.09.033.

Renart, A., Rocha, J. de la, Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K.D. (2010). The Asynchronous State in Cortical Circuits. Science *327*, 587–590. https://doi.org/10.1126/science.1179850.

Repp, B.H., and Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. Journal of Experimental Psychology: Human Perception and Performance *28*, 1085–1099. https://doi.org/10.1037/0096-1523.28.5.1085.

Resulaj, A., Ruediger, S., Olsen, S.R., and Scanziani, M. (2018). First spikes in visual cortex enable perceptual discrimination. ELife *7*, e34044. https://doi.org/10.7554/eLife.34044.

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nat. Neurosci. *2*, 1019–1025. https://doi.org/10.1038/14819.

Ringach, D.L., Shapley, R.M., and Hawken, M.J. (2002a). Orientation selectivity in macaque V1: diversity and laminar dependence. J Neurosci *22*, 5639–5651. https://doi.org/20026567.

Ringach, D.L., Shapley, R.M., and Hawken, M.J. (2002b). Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. J. Neurosci. *22*, 5639–5651. https://doi.org/10.1523/JNEUROSCI.22-13-05639.2002.

Ritchie, J.B., Kaplan, D.M., and Klein, C. (2016). Decoding the Brain: Neural Representation and the Limits of Multivariate Pattern Analysis in Cognitive Neuroscience.

Robie, A.A., Seagraves, K.M., Egnor, S.E.R., and Branson, K. (2017). Machine vision methods for analyzing social interactions. Journal of Experimental Biology *220*, 25–34. https://doi.org/10.1242/jeb.142281.

Robinson, D., and Goldberg, M. (1978). Sensory and Behavioral Properties of Neurons in Posterior Parietal Cortex of Awake, Trained Monkey. Federation Proceedings *37*, 2258–2261. .

Rockland, K.S., and Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. International Journal of Psychophysiology *50*, 19–26. https://doi.org/10.1016/S0167-8760(03)00121-1.

Roelfsema, P.R. (2006). Cortical Algorithms for Perceptual Grouping. Annual Review of Neuroscience *29*, 203–227. https://doi.org/10.1146/annurev.neuro.29.051605.112939.

Roelfsema, P.R., Lamme, V.A.F., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. Nature *395*, 376–381. https://doi.org/10.1038/26475.

Rohe, T., and Noppeney, U. (2015). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. PLOS Biology *13*, e1002073. https://doi.org/10.1371/journal.pbio.1002073.

Rombouts, J.O., Bohte, S.M., Martinez-Trujillo, J., and Roelfsema, P.R. (2015). A learning rule that explains how rewards teach attention. Visual Cognition *23*, 179–205. https://doi.org/10.1080/13506285.2015.1010462.

Romo, R., and Rossi-Pool, R. (2020). Turning Touch into Perception. Neuron *105*, 16–33. https://doi.org/10.1016/j.neuron.2019.11.033.

Rosenberg, M., Zhang, T., Perona, P., and Meister, M. (2021). Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. ELife *10*, e66175. https://doi.org/10.7554/eLife.66175.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., and Foxe, J.J. (2007). Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. Cerebral Cortex *17*, 1147–1153. https://doi.org/10.1093/cercor/bhl024.

Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., et al. (2016). Spike sorting for large, dense electrode arrays. Nat Neurosci *19*, 634–641. https://doi.org/10.1038/nn.4268.

Rossi-Pool, R., Salinas, E., Zainos, A., Alvarez, M., Vergara, J., Parga, N., and Romo, R. (2016). Emergence of an abstract categorical code enabling the discrimination of temporally structured tactile stimuli. PNAS *113*, E7966–E7975. https://doi.org/10.1073/pnas.1618196113.

Rossi-Pool, R., Zainos, A., Alvarez, M., Diaz-deLeon, G., and Romo, R. (2021). A continuum of invariant sensory and behavioral-context perceptual coding in secondary somatosensory cortex. Nat Commun *12*, 2000. https://doi.org/10.1038/s41467-021-22321-x.

Roth, M.M., Dahmen, J.C., Muir, D.R., Imhof, F., Martini, F.J., and Hofer, S.B. (2016). Thalamic nuclei convey diverse contextual information to layer 1 of visual cortex. Nat Neurosci *19*, 299–307. https://doi.org/10.1038/nn.4197.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. Psychonomic Bulletin & Review *16*, 225–237. https://doi.org/10.3758/PBR.16.2.225.

Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. Developmental Neurobiology *71*, 45–61. https://doi.org/10.1002/dneu.20853.

Runyan, C.A., Piasini, E., Panzeri, S., and Harvey, C.D. (2017). Distinct timescales of population coding across cortex. Nature *548*, 92–96. https://doi.org/10.1038/nature23020.

Sachdev, R.N.S., Ebner, F.F., and Wilson, C.J. (2004). Effect of Subthreshold Up and Down States on the Whisker-Evoked Response in Somatosensory Cortex. Journal of Neurophysiology *92*, 3511–3521. https://doi.org/10.1152/jn.00347.2004.

Sachidhanandam, S., Sreenivasan, V., Kyriakatos, A., Kremer, Y., and Petersen, C.C.H. (2013). Membrane potential correlates of sensory perception in mouse barrel cortex. Nat. Neurosci. *16*, 1671–1677. https://doi.org/10.1038/nn.3532.

Sakata, S. (2016). State-dependent and cell type-specific temporal processing in auditory thalamocortical circuit. Scientific Reports *6*, 18873. https://doi.org/10.1038/srep18873.

Sakata, S., and Harris, K.D. (2009). Laminar Structure of Spontaneous and Sensory-Evoked Population Activity in Auditory Cortex. Neuron *64*, 404–418. https://doi.org/10.1016/j.neuron.2009.09.020.

Salkoff, D.B., Zagha, E., McCarthy, E., and McCormick, D.A. (2020). Movement and Performance Explain Widespread Cortical Activity in a Visual Detection Task. Cereb Cortex *30*, 421–437. https://doi.org/10.1093/cercor/bhz206.

Sanders, R.D., Tononi, G., Laureys, S., and Sleigh, J.W. (2012). Unresponsiveness ≠ unconsciousness. Anesthesiology *116*, 946–959. https://doi.org/10.1097/ALN.0b013e318249d0a7.

Schmitt, L.I., Wimmer, R.D., Nakajima, M., Happ, M., Mofakham, S., and Halassa, M.M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. Nature *545*, 219–223. https://doi.org/10.1038/nature22073.

Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K., and Redish, A.D. (2005). Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience *131*, 1–11. https://doi.org/10.1016/j.neuroscience.2004.09.066.

Schnabel, U.H., Bossens, C., Lorteije, J.A.M., Self, M.W., Op de Beeck, H., and Roelfsema, P.R. (2018). Figure-ground perception in the awake mouse and neuronal activity elicited by figure-ground stimuli in primary visual cortex. Scientific Reports *8*. https://doi.org/10.1038/s41598-018-36087-8.

Schneider, D.M., and Mooney, R. (2018). How Movement Modulates Hearing. Annual Review of Neuroscience *41*, 553–572. https://doi.org/10.1146/annurev-neuro-072116-031215.

Schneider, D.M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature *513*, 189–194. https://doi.org/10.1038/nature13724.

Scholl, B., Gao, X., and Wehr, M. (2010). Nonoverlapping Sets of Synapses Drive On Responses and Off Responses in Auditory Cortex. Neuron *65*, 412–421. https://doi.org/10.1016/j.neuron.2010.01.020.

Schölvinck, M.L., Saleem, A.B., Benucci, A., Harris, K.D., and Carandini, M. (2015). Cortical State Determines Global Variability and Correlations in Visual Cortex. J. Neurosci. *35*, 170–178. https://doi.org/10.1523/JNEUROSCI.4994-13.2015.

Schoups, A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. Nature *412*, 549–553. https://doi.org/10.1038/35087601.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. Trends in Cognitive Sciences *12*, 106–113. https://doi.org/10.1016/j.tics.2008.01.002.

Schröter, M.S., Spoormaker, V.I., Schorer, A., Wohlschläger, A., Czisch, M., Kochs, E.F., Zimmer, C., Hemmer, B., Schneider, G., Jordan, D., et al. (2012). Spatiotemporal Reconfiguration of Large-Scale Brain Functional Networks during Propofol-Induced Loss of Consciousness. J. Neurosci. *32*, 12832–12840. https://doi.org/10.1523/JNEUROSCI.6046-11.2012.

Schuman, B., Dellal, S., Prönneke, A., Machold, R., and Rudy, B. (2021). Neocortical Layer 1: An Elegant Solution to Top-Down and Bottom-Up Integration. Annual Review of Neuroscience *44*, 221–252. https://doi.org/10.1146/annurev-neuro-100520-012117.

Seitz, A.R., and Watanabe, T. (2009). The phenomenon of task-irrelevant perceptual learning. Vision Research *49*, 2604–2610. https://doi.org/10.1016/j.visres.2009.08.003.

Sekuler, R., Sekuler, A.B., and Lau, R. (1997). Sound alters visual motion perception. Nature *385*, 308–308. https://doi.org/10.1038/385308a0.

Sela, Y., Vyazovskiy, V.V., Cirelli, C., Tononi, G., and Nir, Y. (2016). Responses in Rat Core Auditory Cortex are Preserved during Sleep Spindle Oscillations. Sleep *39*, 1069–1082. https://doi.org/10.5665/sleep.5758.

Senkowski, D., Saint-Amour, D., Höfle, M., and Foxe, J.J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. NeuroImage *56*, 2200–2208. https://doi.org/10.1016/j.neuroimage.2011.03.075.

Senzai, Y., Fernandez-Ruiz, A., and Buzsáki, G. (2019). Layer-Specific Physiological Features and Interlaminar Interactions in the Primary Visual Cortex of the Mouse. Neuron https://doi.org/10.1016/j.neuron.2018.12.009.

Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci U S A *104*, 6424–6429. https://doi.org/10.1073/pnas.0700622104.

Shadi, K., Dyer, E., and Dovrolis, C. (2020). Multisensory integration in the mouse cortical connectome using a network diffusion model. Network Neuroscience *4*, 1030–1054. https://doi.org/10.1162/netn_a_00164.

Shadlen, M.N., and Newsome, W.T. (1998). The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding. J. Neurosci. *18*, 3870–3896. .

Shaheen, L.A., Slee, S.J., and David, S.V. (2021). Task Engagement Improves Neural Discriminability in the Auditory Midbrain of the Marmoset Monkey. J. Neurosci. *41*, 284–297. https://doi.org/10.1523/JNEUROSCI.1112-20.2020.

Shams, L., and Beierholm, U.R. (2010). Causal inference in perception. Trends in Cognitive Sciences *14*, 425–432. https://doi.org/10.1016/j.tics.2010.07.001.

Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions: What you see is what you hear. Nature *408*, 788–788. https://doi.org/10.1038/35048669.

Shen, S., Jiang, X., Scala, F., Fu, J., Fahey, P., Kobak, D., Tan, Z., Reimer, J., Sinz, F., and Tolias, A.S. (2020). Distinct organization of two cortico-cortical feedback pathways. BioRxiv 2020.02.27.968792. https://doi.org/10.1101/2020.02.27.968792.

Shepard, R.N. (1964). Circularity in Judgments of Relative Pitch. The Journal of the Acoustical Society of America *36*, 2346–2353. https://doi.org/10.1121/1.1919362.

Sheppard, J.P., Raposo, D., and Churchland, A.K. (2013). Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. J Vis *13*, 4. https://doi.org/10.1167/13.6.4.

Sherman, S.M. (2016). Thalamus plays a central role in ongoing cortical functioning. Nature Neuroscience *16*, 533–541. https://doi.org/10.1038/nn.4269.

Shimaoka, D., Harris, K.D., and Carandini, M. (2018). Effects of Arousal on Mouse Sensory Cortex Depend on Modality. Cell Reports *22*, 3160–3167. https://doi.org/10.1016/j.celrep.2018.02.092.

Shuler, M.G., and Bear, M.F. (2006). Reward Timing in the Primary Visual Cortex. Science *311*, 1606–1609. https://doi.org/10.1126/science.1123513.

Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science *348*, 1352–1355. https://doi.org/10.1126/science.aab0551.

Siegle, J.H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T.K., Choi, H., Luviano, J.A., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. Nature *592*, 86–92. https://doi.org/10.1038/s41586-020-03171-x.

Sigler, A., Mohajerani, M.H., and Murphy, T.H. (2009). Imaging rapid redistribution of sensory-evoked depolarization through existing cortical pathways after targeted stroke in mice. PNAS *106*, 11759–11764. .

Sikkens, T., Bosman, C.A., and Olcese, U. (2019). The Role of Top-Down Modulation in Shaping Sensory Processing Across Brain States: Implications for Consciousness. Front. Syst. Neurosci. *13*. https://doi.org/10.3389/fnsys.2019.00031.

Silvanto, J., Cowey, A., Lavie, N., and Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. Nat Neurosci *8*, 143–144. https://doi.org/10.1038/nn1379.

Simons, D., and Chabris, C. (1999). Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. Perception *28*, 1059–1074. https://doi.org/10.1068/p281059.

Simons, D.J., and Levin, D.T. (1997). Change blindness. Trends in Cognitive Sciences *1*, 261–267. https://doi.org/10.1016/S1364-6613(97)01080-2.

Smith, S.L., and Häusser, M. (2010). Parallel processing of visual space by neighboring neurons in mouse visual cortex. Nat Neurosci *13*, 1144–1149. https://doi.org/10.1038/nn.2620.

Sohal, V.S., Zhang, F., Yizhar, O., and Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature *459*, 698–702. https://doi.org/10.1038/nature07991.

Solvi, C., Al-Khudhairy, S.G., and Chittka, L. (2020). Bumble bees display cross-modal object recognition between visual and tactile senses. Science *367*, 910–912. https://doi.org/10.1126/science.aay8064.

Sommer, M.A., and Wurtz, R.H. (2008). Visual Perception and Corollary Discharge. Perception *37*, 408–418. https://doi.org/10.1068/p5873.

Song, Y.-H., Kim, J.-H., Jeong, H.-W., Choi, I., Jeong, D., Kim, K., and Lee, S.-H. (2017). A Neural Circuit for Auditory Dominance over Visual Perception. Neuron *93*, 940–954. https://doi.org/10.1016/j.neuron.2017.01.006.

Spence, C. (2015). Multisensory Flavor Perception. Cell *161*, 24–35. https://doi.org/10.1016/j.cell.2015.03.007.

Spinelli, D.N., Starr, A., and Barrett, T.W. (1968). Auditory specificity in unit recordings from cat's visual cortex. Experimental Neurology *22*, 75–84. https://doi.org/10.1016/0014-4886(68)90020-4.

Spongr, V.P., Flood, D.G., Frisina, R.D., and Salvi, R.J. (1997). Quantitative measures of hair cell loss in CBA and C57BL/6 mice throughout their life spans. The Journal of the Acoustical Society of America *101*, 3546–3553. https://doi.org/10.1121/1.418315.

Spratling, M.W. (2017). A review of predictive coding algorithms. Brain and Cognition *112*, 92–97. https://doi.org/10.1016/j.bandc.2015.11.003.

Sridharan, D., Steinmetz, N.A., Moore, T., and Knudsen, E.I. (2014). Distinguishing bias from sensitivity effects in multialternative detection tasks. Journal of Vision *14*, 16–16. https://doi.org/10.1167/14.9.16.

Srinivasan, M.V., Laughlin, S.B., Dubs, A., and Horridge, G.A. (1982). Predictive coding: a fresh view of inhibition in the retina. Proceedings of the Royal Society of London. Series B. Biological Sciences *216*, 427–459. https://doi.org/10.1098/rspb.1982.0085.

Stein, B.E., and Stanford, T.R. (2008). Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci *9*, 255–266. https://doi.org/10.1038/nrn2331.

Steinmetz, N.A., Zatka-Haas, P., Carandini, M., and Harris, K.D. (2019). Distributed coding of choice, action and engagement across the mouse brain. Nature *576*, 266–273. https://doi.org/10.1038/s41586-019-1787-x.

Storm, J.F., Boly, M., Casali, A.G., Massimini, M., Olcese, U., Pennartz, C.M.A., and Wilke, M. (2017). Consciousness Regained: Disentangling Mechanisms, Brain Systems, and Behavioral Responses. J. Neurosci. *37*, 10882–10893. https://doi.org/10.1523/JNEUROSCI.1838-17.2017.

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. Science *364*, 255–255. https://doi.org/10.1126/science.aav7893.

Stujenske, J.M., Spellman, T., and Gordon, J.A. (2015). Modeling the Spatiotemporal Dynamics of Light and Heat Propagation for In Vivo Optogenetics. Cell Reports *12*, 525–534. https://doi.org/10.1016/j.celrep.2015.06.036.

Stüttgen, M.C., Schwarz, C., and Jäkel, F. (2011). Mapping Spikes to Sensations. Front. Neurosci. *5*. https://doi.org/10.3389/fnins.2011.00125.

Sumby, W.H., and Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. The Journal of the Acoustical Society of America *26*, 212–215. https://doi.org/10.1121/1.1907309.

Supèr, H., Spekreijse, H., and Lamme, V.A.F. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nature Neuroscience *4*, 304–310. https://doi.org/10.1038/85170.

Supèr, H., van der Togt, C., Spekreijse, H., and Lamme, V.A.F. (2004). Correspondence of presaccadic activity in the monkey primary visual cortex with saccadic eye movements. Proceedings of the National Academy of Sciences *101*, 3230–3235. https://doi.org/10.1073/pnas.0400433101.

Supp, G.G., Siegel, M., Hipp, J.F., and Engel, A.K. (2011). Cortical hypersynchrony predicts breakdown of sensory processing during loss of consciousness. Curr. Biol. *21*, 1988–1993. https://doi.org/10.1016/j.cub.2011.10.017.

Svoboda, K., and Li, N. (2018). Neural mechanisms of movement planning: motor cortex and beyond. Current Opinion in Neurobiology *49*, 33–41. https://doi.org/10.1016/j.conb.2017.10.023.

Tafazoli, S., Safaai, H., De Franceschi, G., Rosselli, F.B., Vanzella, W., Riggi, M., Buffolo, F., Panzeri, S., and Zoccolan, D. (2017). Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. Elife *6*. https://doi.org/10.7554/eLife.22794.

Tanahira, C., Higo, S., Watanabe, K., Tomioka, R., Ebihara, S., Kaneko, T., and Tamamaki, N. (2009). Parvalbumin neurons in the forebrain as revealed by parvalbumin-Cre transgenic mice. Neuroscience Research *63*, 213–223. https://doi.org/10.1016/j.neures.2008.12.007.

Teufel, C., and Fletcher, P.C. (2020). Forms of prediction in the nervous system. Nature Reviews Neuroscience 1–12. https://doi.org/10.1038/s41583-020-0275-5.

Thorpe, S. (1989). Local vs. Distributed Coding. Intellectica *8*, 3–40. https://doi.org/10.3406/intel.1989.873.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature *381*, 520–522. https://doi.org/10.1038/381520a0.

van der Togt, C., Kalitzin, S., Spekreijse, H., Lamme, V.A.F., and Supèr, H. (2006). Synchrony Dynamics in Monkey V1 Predict Success in Visual Detection. Cereb Cortex *16*, 136–148. https://doi.org/10.1093/cercor/bhi093.

Tohmi, M., Meguro, R., Tsukano, H., Hishida, R., and Shibuki, K. (2014). The Extrageniculate Visual Pathway Generates Distinct Response Properties in the Higher Visual Areas of Mice. Current Biology *24*, 587–597. https://doi.org/10.1016/j.cub.2014.01.061.

Tolhurst, D.J., Movshon, J.A., and Thompson, I.D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. Exp Brain Res *41*, 414–419. https://doi.org/10.1007/BF00238900.

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience *17*, 450–461. https://doi.org/10.1038/nrn.2016.44.

Trainito, C., Nicolai, C. von, Miller, E.K., and Siegel, M. (2019). Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex. Current Biology *29*, 2973-2982.e5. https://doi.org/10.1016/j.cub.2019.07.051.

Tseng, S.-Y., Chettih, S.N., Arlt, C., Barroso-Luque, R., and Harvey, C.D. (2022). Shared and specialized coding across posterior cortical areas for dynamic navigation decisions. Neuron https://doi.org/10.1016/j.neuron.2022.05.012.

Tsodyks, M.V., Skaggs, W.E., Sejnowski, T.J., and McNaughton, B.L. (1997). Paradoxical Effects of External Modulation of Inhibitory Interneurons. J. Neurosci. *17*, 4382–4388. https://doi.org/10.1523/JNEUROSCI.17-11-04382.1997.

Untergehrer, G., Jordan, D., Kochs, E.F., Ilg, R., and Schneider, G. (2014). Fronto-Parietal Connectivity Is a Non-Static Phenomenon with Characteristic Changes during Unconsciousness. PLOS ONE *9*, e87498. https://doi.org/10.1371/journal.pone.0087498.

Vaiceliunaite, A., Erisken, S., Franzen, F., Katzner, S., and Busse, L. (2013). Spatial integration in mouse primary visual cortex. Journal of Neurophysiology *110*, 964–972. https://doi.org/10.1152/jn.00138.2013.

Vaknin, G., DiScenna, P.G., and Teyler, T.J. (1988). A method for calculating current source density (CSD) analysis without resorting to recording sites outside the sampling volume. Journal of Neuroscience Methods *24*, 131–135. https://doi.org/10.1016/0165-0270(88)90056-8.

Vallar, G. (1998). Spatial hemineglect in humans. Trends in Cognitive Sciences *2*, 87–97. https://doi.org/10.1016/S1364-6613(98)01145-0.

Van der Werf, Y.D., Witter, M.P., and Groenewegen, H.J. (2002). The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. Brain Research Reviews *39*, 107–140. https://doi.org/10.1016/S0165-0173(02)00181-9.

van Atteveldt, N., Murray, M.M., Thut, G., and Schroeder, C.E. (2014). Multisensory Integration: Flexible Use of General Operations. Neuron *81*, 1240–1253. https://doi.org/10.1016/j.neuron.2014.02.044.

Vangeneugden, J., van Beest, E.H., Cohen, M.X., Lorteije, J.A.M., Mukherjee, S., Kirchberger, L., Montijn, J.S., Thamizharasu, P., Camillo, D., Levelt, C.N., et al. (2019). Activity in Lateral Visual Areas Contributes to Surround Suppression in Awake Mouse V1. Curr. Biol. *29*, 4268-4275.e7. https://doi.org/10.1016/j.cub.2019.10.037.

Vélez-Fort, M., Bracey, E.F., Keshavarzi, S., Rousseau, C.V., Cossell, L., Lenzi, S.C., Strom, M., and Margrie, T.W. (2018). A Circuit for Integration of Head- and Visual-Motion Signals in Layer 6 of Mouse Primary Visual Cortex. Neuron *98*, 179-191.e6. https://doi.org/10.1016/j.neuron.2018.02.023.

Vergara, J., Rivera, N., Rossi-Pool, R., and Romo, R. (2016). A Neural Parametric Code for Storing Information of More than One Sensory Modality in Working Memory. Neuron *89*, 54–62. https://doi.org/10.1016/j.neuron.2015.11.026.

Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J.A. (2015a). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron *86*, 740–754. https://doi.org/10.1016/j.neuron.2015.03.028.

Vinck, M., Bos, J.J., Van Mourik-Donga, L.A., Oplaat, K.T., Klein, G.A., Jackson, J.C., Gentet, L.J., and Pennartz, C.M.A. (2015b). Cell-Type and State-Dependent Synchronization among Rodent Somatosensory, Visual, Perirhinal Cortex, and Hippocampus CA1. Front Syst Neurosci *9*, 187. https://doi.org/10.3389/fnsys.2015.00187.

Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J.A. (2015c). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. Neuron *86*, 740–754. https://doi.org/10.1016/j.neuron.2015.03.028.

Vinck, M., Bos, J.J., Van Mourik-Donga, L.A., Oplaat, K.T., Klein, G.A., Jackson, J.C., Gentet, L.J., and Pennartz, C.M.A. (2016). Cell-Type and State-Dependent Synchronization among Rodent Somatosensory, Visual, Perirhinal Cortex, and Hippocampus CA1. Front. Syst. Neurosci. *9*. https://doi.org/10.3389/fnsys.2015.00187.

Vugt, B. van, Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P.R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. Science eaar7186. https://doi.org/10.1126/science.aar7186.

Wallace, M.T., Ramachandran, R., and Stein, B.E. (2004). A revised view of sensory cortical parcellation. Proceedings of the National Academy of Sciences of the United States of America *101*, 2167–2172. .

Wander, J.D., Blakely, T., Miller, K.J., Weaver, K.E., Johnson, L.A., Olson, J.D., Fetz, E.E., Rao, R.P.N., and Ojemann, J.G. (2013). Distributed cortical adaptation during learning of a brain–computer interface task. PNAS *110*, 10818–10823. .

Wang, Q., and Burkhalter, A. (2007). Area map of mouse visual cortex. J. Comp. Neurol. *502*, 339–357. https://doi.org/10.1002/cne.21286.

Wang, J.X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J.Z., Hassabis, D., and Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. https://doi.org/10.1101/295964.

Wang, L., McAlonan, K., Goldstein, S., Gerfen, C.R., and Krauzlis, R.J. (2020). A Causal Role for Mouse Superior Colliculus in Visual Perceptual Decision-Making. J. Neurosci. *40*, 3768–3782. https://doi.org/10.1523/JNEUROSCI.2642-19.2020.

Wang, Q., Gao, E., and Burkhalter, A. (2011). Gateways of Ventral and Dorsal Streams in Mouse Visual Cortex. J. Neurosci. *31*, 1905–1918. https://doi.org/10.1523/JNEUROSCI.3488-10.2011.

Wang, Q., Sporns, O., and Burkhalter, A. (2012). Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. J Neurosci *32*, 4386–4399. https://doi.org/10.1523/JNEUROSCI.6063-11.2012.

Wang, X., Chen, C., Zhang, D., and Yao, H. (2014). Cumulative latency advance underlies fast visual processing in desynchronized brain state. PNAS *111*, 515–520. https://doi.org/10.1073/pnas.1316166111.

Wang, Y., Celebrini, S., Trotter, Y., and Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: Electrophysiological evidence. BMC Neurosci *9*, 79. https://doi.org/10.1186/1471-2202-9-79.

Waskom, M.L., Okazawa, G., and Kiani, R. (2019). Designing and Interpreting Psychophysical Investigations of Cognition. Neuron *104*, 100–112. https://doi.org/10.1016/j.neuron.2019.09.016.

Weinberger, N.M., Javid, R., and Lepan, B. (1993). Long-term retention of learning-induced receptive-field plasticity in the auditory cortex. PNAS *90*, 2394–2398. https://doi.org/10.1073/pnas.90.6.2394.

Whitlock, J.R., Sutherland, R.J., Witter, M.P., Moser, M.-B., and Moser, E.I. (2008). Navigating from hippocampus to parietal cortex. PNAS *105*, 14755–14762. https://doi.org/10.1073/pnas.0804216105.

Whitlock, J.R., Pfuhl, G., Dagslott, N., Moser, M.-B., and Moser, E.I. (2012). Functional Split between Parietal and Entorhinal Cortices in the Rat. Neuron *73*, 789–802. https://doi.org/10.1016/j.neuron.2011.12.028.

Wilber, A.A., Clark, B.J., Demecha, A.J., Mesina, L., Vos, J.M., and McNaughton, B.L. (2015). Cortical connectivity maps reveal anatomically distinct areas in the parietal cortex of the rat. Front. Neural Circuits *8*. https://doi.org/10.3389/fncir.2014.00146.

Wilber, A.A., Skelin, I., Wu, W., and McNaughton, B.L. (2017). Laminar Organization of Encoding and Memory Reactivation in the Parietal Cortex. Neuron *95*, 1406-1419.e5. https://doi.org/10.1016/j.neuron.2017.08.033.

Williams, A.H., Kim, T.H., Wang, F., Vyas, S., Ryu, S.I., Shenoy, K.V., Schnitzer, M., Kolda, T.G., and Ganguli, S. (2018). Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. Neuron *98*, 1099-1115.e8. https://doi.org/10.1016/j.neuron.2018.05.015.

Williams, A.M., Angeloni, C.F., and Geffen, M.N. (2021). Sound improves neuronal encoding of visual stimuli in mouse primary visual cortex. BioRxiv 2021.08.03.454738. https://doi.org/10.1101/2021.08.03.454738.

Wilson, N.R., Runyan, C.A., Wang, F.L., and Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. Nature *488*, 343–348. https://doi.org/10.1038/nature11347.

Wimmer, R.D., Schmitt, L.I., Davidson, T.J., Nakajima, M., Deisseroth, K., and Halassa, M.M. (2015). Thalamic control of sensory selection in divided attention. Nature *526*, 705–709. https://doi.org/10.1038/nature15398.

Wörgötter, F., Suder, K., Zhao, Y., Kerscher, N., Eysel, U.T., and Funke, K. (1998). State-dependent receptive-field restructuring in the visual cortex. Nature *396*, 165–168. https://doi.org/10.1038/24157.

Wurtz, R.H. (1968). Visual Cortex Neurons: Response to Stimuli during Rapid Eye Movements. Science *162*, 1148–1150. https://doi.org/10.1126/science.162.3858.1148.

Xue, M., Atallah, B.V., and Scanziani, M. (2014). Equalizing excitation-inhibition ratios across visual cortical neurons. Nature *511*, 596–600. https://doi.org/10.1038/nature13321.

Yang, T., and Maunsell, J.H.R. (2004). The Effect of Perceptual Learning on Neuronal Responses in Monkey Visual Area V4. J. Neurosci. *24*, 1617–1626. https://doi.org/10.1523/JNEUROSCI.4442-03.2004.

Yang, G.R., Joglekar, M.R., Song, H.F., Newsome, W.T., and Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. Nature Neuroscience *22*, 297–306. https://doi.org/10.1038/s41593-018-0310-2.

Yeomans, J.S., and Frankland, P.W. (1995). The acoustic startle reflex: neurons and connections. Brain Research Reviews *3*, 301–314. .

Yin, P., Strait, D.L., Radtke-Schuller, S., Fritz, J.B., and Shamma, S.A. (2020). Dynamics and Hierarchical Encoding of Non-compact Acoustic Categories in Auditory and Frontal Cortex. Current Biology *30*, 1649-1663.e5. https://doi.org/10.1016/j.cub.2020.02.047.

Young, H., Belbut, B., Baeta, M., and Petreanu, L. (2021). Laminar-specific cortico-cortical loops in mouse visual cortex. ELife *10*, e59551. https://doi.org/10.7554/eLife.59551.

Zagha, E., Casale, A.E., Sachdev, R.N.S., McGinley, M.J., and McCormick, D.A. (2013). Motor Cortex Feedback Influences Sensory Processing by Modulating Network State. Neuron *79*, 567–578. https://doi.org/10.1016/j.neuron.2013.06.008.

Zagha, E., Erlich, J.C., Lee, S., Lur, G., O'Connor, D.H., Steinmetz, N.A., Stringer, C., and Yang, H. (2022). The importance of accounting for movement when relating neuronal activity to sensory and cognitive processes. J. Neurosci. https://doi.org/10.1523/JNEUROSCI.1919-21.2021.

Zatka-Haas, P., Steinmetz, N.A., Carandini, M., and Harris, K.D. (2021). Sensory coding and the causal impact of mouse cortex in a visual decision. ELife *10*, e63163. https://doi.org/10.7554/eLife.63163.

Zhang, S., Xu, M., Kamigaki, T., Do, J.P.H., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Long-range and local circuits for top-down modulation of visual cortex processing. Science *345*, 660–665. https://doi.org/10.1126/science.1254126.

Zhang, S., Xu, M., Chang, W.-C., Ma, C., Hoang Do, J.P., Jeong, D., Lei, T., Fan, J.L., and Dan, Y. (2016). Organization of long-range inputs and outputs of frontal cortex for top-down control. Nat Neurosci *19*, 1733–1742. https://doi.org/10.1038/nn.4417.

Zhong, L., Zhang, Y., Duan, C.A., Deng, J., Pan, J., and Xu, N. (2019). Causal contributions of parietal cortex to perceptual decision-making during stimulus categorization. Nat Neurosci *22*, 963–973. https://doi.org/10.1038/s41593-019-0383-6.

Zingg, B., Hintiryan, H., Gou, L., Song, M.Y., Bay, M., Bienkowski, M.S., Foster, N.N., Yamashita, S., Bowman, I., Toga, A.W., et al. (2014). Neural Networks of the Mouse Neocortex. Cell *156*, 1096–1111. https://doi.org/10.1016/j.cell.2014.02.023.

Zmarz, P., and Keller, G.B. (2016). Mismatch Receptive Fields in Mouse Visual Cortex. Neuron *92*, 766–772. https://doi.org/10.1016/j.neuron.2016.09.057.

Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. Nature *497*, 482–485. https://doi.org/10.1038/nature12077.

Zohary, E., Shadlen, M.N., and Newsome, W.T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. Nature *370*, 140–143. https://doi.org/10.1038/370140a0.

List of publications

Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., **Oude Lohuis, M.N.**, and Mainen, Z.F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. Nat. Commun. 9, 1000. https://doi.org/10.1038/s41467-018-03438-y.

Olcese, U., **Oude Lohuis, M.N**., and Pennartz, C.M.A. (2018). Sensory Processing Across Conscious and Nonconscious Brain States: From Single Neurons to Distributed Networks for Inferential Representation. Front. Syst. Neurosci. 12 (49). https://doi.org/10.3389/fnsys.2018.00049.

Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., **Oude Lohuis, M.N**., and Mainen, Z.F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. Neuron 106:166-176e6. https://doi.org/10.1016/j.neuron.2020.01.017.

Oude Lohuis, M.N., Canton, A.C., Pennartz, C.M.A., and Olcese, U. (2021). Higher Order Visual Areas Enhance Stimulus Responsiveness in Mouse Primary Visual Cortex. Cereb. Cortex. https://doi.org/10.1093/cercor/bhab414.

Oude Lohuis, M.N., Pie, J.L., Marchesi, P., Montijn, J.S., de Kock, C.P.J., Pennartz, C.M.A., and Olcese, U. (2022a). Multisensory task demands temporally extend the causal requirement for visual cortex in perception. Nat. Commun. 13, 2864. https://doi.org/10.1038/s41467-022-30600-4.

Oude Lohuis, M.N., Marchesi, P., Pennartz, C.M.A., and Olcese, U. (2022b). Functional (ir)relevance of posterior parietal cortex during audiovisual change detection. J. Neurosci. 42, 5229–5245 https://doi.org/10.1523/JNEUROSCI.2150-21.2022.

Mertens, P.E., Marchesi, P., **Oude Lohuis, M.N.**, Krijger, Q., Pennartz, C.M., and Lansink, C.S. (2021). Coherent mapping of position and head direction across auditory and visual cortex. BioRxiv. https://doi.org/10.1101/2021.07.30.452931.

Oude Lohuis, M.N., Marchesi, P., Olcese, U., and Pennartz, C.M.A. (2022). Triple dissociation of visual, auditory and motor processing in primary visual cortex. BioRxiv. https://doi.org/10.1101/2022.06.29.498156.

Summary

Summary of: "Contextual signals in visual cortex: how sounds, state, and task setting shape how we see"

What we see is not always what we get. Even though the light that hits the retina might convey the same images, how visual information is processed and what we eventually do with it depend on many contextual factors such as our internal state, movements, other senses and any task we are performing. In this thesis, I studied how the sensory processing of the same visual input in the visual cortex of mice is affected by different contextual factors. In **Chapter 2** we describe how activity in the primary visual cortex (V1) is affected by recurrent activity originating within higher visual areas. This recurrent activity specifically enhanced stimulus responsiveness of V1 neurons by more strongly amplifying weaker compared to stronger sensory-evoked responses and in a similar manner during anesthetized as awake conditions. In **Chapter 3**, we found that visual stimuli evoked similar early sensory-evoked activity in V1, but that later activity strongly depended on whether mice were trained to report the visual stimuli, and on the specific task. Specifically, adding a second modality to the task demands extended the temporal window during which V1 was causally involved in visual perception. In **Chapter 4**, we reported that not only visual stimuli but also sounds led to strong activity increases in V1, and that this response was composed of two distinct components. We found a transient auditory-related component originating from auditory cortex, but also a motor-related component reflecting a rapid effect of sound-evoked movements on V1 activity. Finally, in **Chapter 5** we studied the role of Posterior Parietal Cortex (PPC) in an audiovisual change detection task. Despite extensive single-neuron and population-level encoding of task-relevant visual and auditory stimuli, as well as upcoming behavioral responses, optogenetic inactivation of PPC did not affect task performance. Whereas these contextual factors have previously been studied in isolation, we now have an integrative view of their interactions with early visual processing. Feedback from higher visual areas targets visually-responsive neurons, while auditory and motor signals are present in distinct subsets of neurons giving rise to a diverse palette of contextual channels to the early visual system. In conclusion, after this thesis we gained a better understanding of how the visual system flexibly operates and of how factors beyond visual information determine what we actually see.

Nederlandse samenvatting

Samenvatting van: "Contextuele signalen in de visuele cortex: hoe geluid, onze staat, en de taak beïnvloeden hoe we zien"

Wat we zien hangt niet alleen af van het licht dat op onze retina valt. Visuele verwerking wordt mede bepaald door onze verwachtingen, of er belangrijke dingen zijn om op te letten en wat we met onze andere zintuigen waarnemen. Het is bekend dat deze contextuele factoren een groot effect hebben op latere stadia van de visuele verwerking, in de hogere gebieden van de cortex. Deze thesis beschrijft een aantal experimenten waarin gekeken werd hoe de activiteit en functie van de primaire visuele cortex in muizen (een vroeg stadium van visuele verwerking) afhangt van contextuele factoren.
Uit dit onderzoek blijkt dat visuele verwerking ook in vroege stadia geen statisch proces is. In **Hoofdstuk 2** onderzochten we hoe hogere visuele gebieden invloed hebben op de verwerking in lagere gebieden binnen het visuele systeem, en vinden dat deze specifiek zwakke responsen kunnen versterken. In **Hoofdstuk 3**, rapporteren we dat de tijdspanne dat de primaire visuele cortex nodig is in een detectie-taak afhangt van de taak: als het dier gelijktijdig een ander zintuig (geluid of tast) moet monitoren duurt de causale betrokkenheid van de visuele cortex langer. De taak had ook effect op de elektrische activiteit. Kort na de stimulus was de activiteit van neuronen in de visuele cortex tamelijk onveranderlijk, onafhankelijk van de taak. Maar op een later tijdspunt hing de activiteit sterk af van het (gedragsmatig) detecteren van de stimulus en vervolgens het rapporteren ervan, maar ook van de motorische bewegingen en verhoogde alertheid. Ook geluiden hadden sterke effecten op de primaire visuele cortex **(Hoofdstuk 4)**. Omdat geluiden echter ook vaak bewegingen rond de snuit veroorzaakten, en deze weer effecten hadden op visuele cortex, had geluid effect op de visuele cortex via twee processen. Auditieve signalen en motor signalen waren te vinden in verschillende neuronen in de primaire visuele cortex en lieten visuele verwerking intact. Ten slotte beschrijft dit proefschrift in **Hoofdstuk 5** hoe hogere gebieden zoals de pariëtale cortex ook al deze signalen kunnen laten zien die samenhangen met een audiovisuele detectietaak, maar toch overbodig zijn voor het uitvoeren van de taak. Deze thesis heeft ons kennis gebracht hoe geluid, beweging, taak, en zicht interacteren in een georganiseerde manier en een flexibele context verschaffen voor vroege stadia van visuele verwerking.

Acknowledgements

This work was supported on many fronts by a host of different people.

Umberto Olcese, my sincerest gratitude goes to your dedication to my trajectory. I've learned from you across the whole range of scientific skills: from pulling a glass micropipette to scientific story writing, from student supervision to project management. It was ever a pleasure working with you thanks to your availability, eye for detail, and diverse interests. Then I'd like to thank my promotor **Cyriel Pennartz**. Your ideas on novel research projects and assessment of *which* science really contributes to further knowledge was critical. Our discussions and papers were decidedly sharpened by crosslinking to other fields and older lines of research.

There are many other scientific colleagues who provided valuable contributions. **Pietro Marchesi**, we managed to collaborate on three out of four chapters and I regret that you are not on all of them. Thanks for the insightful discussions and segue into population dynamics. **Jean Pie**, I am so glad we were together on our journey of the 'second bump'. It was an immense joy to ride the camel together, calibrating our methods, concepts and problems. Thanks for your passionate engagement. Further thanks for exploration and guidance on this project by **Jorrit Montijn**, and **Chris de Kock**.

Gratitude goes out to students who joined at various stages of the project. **Adrianna Mantzafou**, it was a pleasure starting to look with you at aroused mouse pupils, and with **Klara Gawor** on how neural activity across the brain is different in aroused mice. **Alexis Cerván Canton**, thanks for your valuable contributions during early stages of higher order feedback effects on V1 (Chapter 2). Thanks to all three of you for help with behavioral training, and I really had a good time with each of you. I thank **Sungho Hong** for training in computational neuroscience in Okinawa, Japan, and **Román Rossi-Pool** for insightful discussions and collaborative work during a lab exchange in Mexico that got abruptly cut short due to a certain virus. I thank people from the **hardware platform** for valuable contributions to the experimental setups used for these investigations and **Chris**, **Rob**, and **Miriam** for taking care of the animals and ethical design.

It was great to start this PhD journey together with **Jean** and **Julien**, encountering and sharing the same shitty issues at the same time. Julien, you really bring out the human side of every colleague and Jean you just always bring every colleague. I also learned a lot from PhD colleagues at sister institutes, **Bastijn, Koen, Nora, Doris, Sybren, Huub, Leonie,** and **Esther**. We had such good times at conferences, I learned a lot about other fields of neuroscience, and through our informal lab meetings it became clear that issues were not specific to my project or our lab, but shared, and this made them somehow less bad (and worse if they involved structural problems in science).

Guido, Roos, Tom, Sven, Lianne, Angelica, Eric, João, Jeroen, Judith, Mariel, Mathis, Medina, Lilian, Paul, Reinder: it was amazing to have you around with such a good atmosphere, distraction, support and discussion. I hope Friday afternoons remain characterized by the choice which of the three places to get a beer, that lab meetings regretfully stay on Mnday mornings at 9:00 AM and that the cup consumption of Pietro is going down as fast as the diversity of Reinder's bread topping is going up.

Thanks to everybody at SILS-CNS for fun times: **Sylvie, Kitty, Elise, Rico, Cindy, Diana, Niek, Reinofke, Elias,** and **Swip.** I'd also like to thank the members of my **PhD committee** for putting time and effort in reading this thesis. Creative credits go to **Diogo Matias** for the amazing cover art. Finally, I'd like to thank my girlfriend **Marianne** for all her support.

Matthijs

P.s. I am sorry if I forgot to accredit someone.

P.p.s. Lastly, I want to sincerely apologize for an all-male author list. Of course the composition of contributors to individual projects is subject to chance, but we still have a long way to go towards equal opportunities for everybody in science.