How many neurons are sufficient for perception of cortical activity?

Henry W. P. Dalgleish*[1,3], Lloyd E. Russell*[1], Adam M. Packer*[1,2], Arnd Roth[1], Oliver M. Gauld[1], Francesca Greenstreet[1,3], Emmett J. Thompson[1,3], Michael Häusser[1]

[1] *Wolfson Institute for Biomedical Research, University College London, Gower Street, London WC1E 6BT, UK*

[2] *Present address: Department of Physiology, Anatomy and Genetics, University of Oxford, UK*

[3] *Present address: UCL Sainsbury Wellcome Centre for Neural Circuits and Behaviour, London, UK*

* These authors contributed equally

Correspondence should be addressed to M.H. (m.hausser@ucl.ac.uk)

Abstract

Many theories of brain function assume that information is encoded and behaviour is controlled through sparse, distributed patterns of activity. It is therefore crucial to place a lower bound on the amount of neural activity that can drive behaviour and to understand how neuronal networks operate within these constraints. We use an all-optical approach to test this lower limit by driving behaviour with targeted two-photon optogenetic activation of small ensembles of L2/3 pyramidal neurons in mouse barrel cortex while using two-photon calcium imaging to record the impact on the local network. By precisely titrating the number of neurons in activated ensembles we demonstrate that the lower bound for detection of cortical activity is ~14 pyramidal neurons. We show that there is a very steep sigmoidal relationship between the number of activated neurons and behavioural output, saturating at only ~37 neurons, and that this relationship can shift with learning. By simultaneously measuring activity in the local network, we show that the activation of stimulated ensembles is balanced by the suppression of neighbouring neurons. This surprising behavioural sensitivity in the face of potent network suppression supports the sparse coding hypothesis and suggests that perception of cortical activity balances a trade-off between minimizing the impact of noise while efficiently detecting relevant signals.

Introduction

How does activity in neural circuits give rise to behaviour? While mammalian brains are composed of many millions (Herculano-Houzel, Mota, & Lent, 2006) or billions (Herculano-Houzel, 2009; Herculano-Houzel, Collins, Wong, & Kaas, 2007) of neurons it has long been postulated that they operate – encoding information and controlling behaviour – through the activity of small subsets of those neurons (Barlow, 1972). Indeed, the hypothesis that brains use sparse, distributed activity patterns is supported computationally (Kanerva, 1993; Olshausen & Field, 1996), energetically (Attwell & Laughlin, 2001; Lennie, 2003; Schölvinck, Howarth, & Attwell, 2008) and experimentally (Barth & Poulet, 2012; Olshausen & Field, 2004; Wolfe, Houweling, & Brecht, 2010). A major factor thought to govern such sparse coding is neuronal inhibition (Haider & McCormick, 2009; Isaacson & Scanziani, 2011) which serves to balance and control recurrent excitation (Denève & Machens, 2016; Haider, Häusser, & Carandini, 2013; Murphy & Miller, 2009; Packer & Yuste, 2011; Pehlevan & Sompolinsky, 2014; Sadeh & Clopath, 2020; Tsodyks, Skaggs, Sejnowski, & Mcnaughton, 1997; van Vreeswijk & Sompolinsky, 1996; Wehr & Zador, 2003; Wolf, Engelken, Puelma-Touzel, Weidinger, & Neef, 2014) and shape neuronal output (Borg-Graham, Monier, & Frégnac, 1998; Cardin, Kumbhani, Contreras, & Palmer, 2010; Isaacson & Scanziani, 2011; Lee et al., 2012; Wilson, Runyan, Wang, & Sur, 2012). Two key questions are therefore: (1) What is the lower bound of activity that can be behaviourally salient? and (2) How does such activity interact with the local network?

Classical microstimulation experiments have demonstrated that focal activation of cortical regions can influence decision-making, providing a direct causal link between neural activity and behaviour (Cohen & Newsome, 2004; Murasugi, Salzman, & Newsome, 1993; Salzman, Britten, & Newsome, 1990; Salzman, Murasugi, Britten, & Newsome, 1992). This landmark work has been complemented by more recent studies showing that optogenetic activation of dozens to hundreds of cortical neurons can be directly detected (Huber et al., 2008; Histed and Maunsell, 2014). A further refinement of this approach was provided by patch-clamp recording, which revealed that strong electrical stimulation of even a single neuron can be detected in a cell-type and spike timing-dependent manner (Houweling and Brecht, 2008; Doron et al., 2014), to the extent that they can modulate behaviour in sensory-guided tasks (Tanke, Borst and Houweling, 2018). While these studies have provided important approximations of the numbers of neurons required to trigger and manipulate behaviour, they suffered from several important limitations. Firstly, electrical stimulation techniques are ill-suited to titrating the number of neurons stimulated and offer limited targeting specificity, either indiscriminately activating swathes of cortex or activating single neurons. Secondly, one-photon optogenetic approaches, while limiting direct excitation to genetically defined neurons, only offer post-hoc estimation of the number of neurons stimulated from histology (Huber et al., 2008). Finally, in none of these studies has it been possible to carefully assess the impact of the stimulation on the local network, an essential step if we are to understand the link between activity generated by stimulation and behaviour.

A parallel body of work has investigated the influence of neural activity on local networks *in vivo*, demonstrating that small numbers of active neurons can have a large impact on local network dynamics and brain state. Strong stimulation of single pyramidal neurons in L2/3 has been shown to recruit ~2% of local excitatory neurons and ~30% of local inhibitory neurons (Kwan and Dan, 2012). A single pyramidal neuron spike in L5 is estimated to recruit ~28 post-synaptic neurons (London et al., 2010) and in L2/3 can trigger strong disynaptic inhibition (Jouhanneau, Kremkow and Poulet, 2018). Single neurons can also trigger global switches in brain state (Li, Poo and Dan, 2009), influence network synchronization (Bonifazi et al., 2009) and have a direct impact on motor output (Brecht et al., 2004). However, work investigating the impact of sparse activation on the local network *in vivo* has largely been done under anaesthesia (London et al., 2010; Kwan and Dan, 2012; Jouhanneau, Kremkow and Poulet, 2018) which influences state-dependent cortical processing (Niell and Stryker, 2010; Crochet et al., 2011; Harris and Thiele, 2011) and prevents the study of behaviour.

Combining simultaneous targeted stimulation with readout of effects on the local network during behaviour will allow us to define the local network input-output function. This will yield better understanding of neural network operation, analogously to how measuring single-neuron input-output functions has transformed our understanding of information processing in single neurons (Magee, 2000; Poirazi, Brannon and Mel, 2003; London et al., 2010; Major, Larkum and Schiller, 2013). Moreover, it will allow us to determine how this network input-output function in turn influences the psychometric sensitivity to neural activity, which theoretical work

predicts is crucial for understanding the link between neural circuit activity and behaviour (Bernardi, Doron, Brecht, & Lindner, 2020; Bernardi & Lindner, 2017, 2019; Cai et al., 2020). While some studies combining readout with manipulation have made significant progress in this direction (Ceballo, Bourg, et al., 2019; Ceballo, Piwkowska, Bourg, Daret, & Bathellier, 2019; Salzman et al., 1990; Znamenskiy & Zador, 2013) they have lacked spatial resolution and targeting flexibility either on the level of readout or stimulation. Measuring network input-output functions at cellular resolution during perception is likely to yield pivotal insights into how neural populations generate behaviour.

Here we have activated ensembles of varying numbers of neurons in L2/3 barrel cortex of awake mice trained to detect direct cortical photostimulation. We took advantage of recently developed all-optical approaches combining two-photon calcium imaging and two-photon optogenetics with digital holography to allow us to activate specifically targeted ensembles of neurons while simultaneously recording the response of the local network (Carrillo-Reid et al., 2019; Carrillo-Reid, Yang, Bando, Peterka, & Yuste, 2016; Emiliani, Cohen, Deisseroth, & Hausser, 2015; Mardinly et al., 2018; Marshel et al., 2019; Packer, Russell, Dalgleish, & Hausser, 2015; Russell et al., 2019; Shemesh et al., 2017). We combined this all-optical approach with an operant conditioning paradigm in which animals were required to report the targeted two-photon activation of arbitrary ensembles of pyramidal neurons in L2/3 barrel cortex to gain rewards. In trained animals we investigated how behaviour and network response vary as a function of the ensemble size stimulated. We show that animals are sensitive to the activation of surprisingly small numbers of

neurons (~14) and demonstrate that activating roughly double this number of neurons (~37) is sufficient for detection to saturate. Moreover we show that this perceptual threshold is plastic, and decreases with learning. We also demonstrate that while detection rates increase with increasing stimulation, the surrounding network responds with matched suppression which maintains the balance of activation and suppression at a level consistent with spontaneous epochs.

Results

Targeted two-photon optogenetic activation of neural ensembles in L2/3 barrel cortex can drive behaviour

To investigate both the behavioural and network response to precisely controlled levels of cortical excitation we combined an upgraded version of our previously reported all-optical setup (incorporating 3D volumetric imaging using an ETL, a more powerful two-photon (2P) photostimulation laser and an additional light-path for one-photon (1P) photostimulation; see *Methods,* Packer *et al.*, 2015) with an operant conditioning paradigm whereby mice are trained to report the activation of excitatory neurons in barrel cortex with either 1P or 2P optogenetic stimulation (**Figure 1**). We expressed the calcium indicator GCaMP6s and the two-photon activatable somatically-restricted opsin C1V1 in neurons of L2/3 barrel cortex (**Figure 1a**: inset right, **Figure 1 – figure supplement 1**, see *Methods*). This strategy allows us to flexibly activate specific ensembles of neurons in a given cortical population (**Figure 1a**: Pixel STA, **Figure 1 – figure supplement 2**) with high spatial resolution (**Figure 1 – figure supplement 2a**: HWHM 5 μm laterally, 20 μm axially) while performing simultaneous two-photon calcium imaging. Inspired by previous work (Histed & Maunsell, 2014; Huber et al., 2008) we devised a training paradigm in which animals were conditioned to detect bulk activation of barrel cortex via 1P photostimulation. After task acquisition we progressively lowered the 1P stimulation intensity (which reduces the number, reliability and spatial extent of activated neurons (see *Methods* for photostimulation details)), before transitioning animals to detect 2P photostimulation of specific groups of neurons (**Figure 1b**, **Figure 1 – figure supplement 3**). We used a simple un-cued go/catch trial design where pseudo-

randomly interleaved go trials (1P or 2P photostimulation) or catch trials (no photostimulation) were delivered after animals successfully withheld licking for a variable period (**Figure 1c**: top, see *Methods*). On go trials the presence/absence of licks in the post-stimulus response window was scored as hits/misses, with hits triggering delivery of a sucrose reward (**Figure 1c**: bottom). On catch trials the presence/absence of licks were scored as false alarms/correct rejects and neither outcome was punished or rewarded.

Animals underwent 3 training phases (**Figure 1c**: bottom, **Figure 1 – figure supplement 3a**), beginning with Phase 1 where 1P go trials and catch trials were interleaved in equal proportions. In the first training session (**Figure 1 – figure supplement 3b**), go trials of 10 mW 1P photostimuli were delivered and were automatically rewarded irrespective of the animal's response (**Figure 1 – figure supplement 3b**, blue line). Animals readily learned to detect photostimulation as shown by an increase in proactive lick responses post-stimulus but before the automatic reward (delivered at 0.5 s post-stimulus onset) and decreases in reaction time mean and s.d. across the first 20 – 100 trials (**Figure 1 – figure supplement 3c - e**), often within the first session (**Figure 1 – figure supplement 3b**). Once animals showed evidence of learning, automatic rewards were turned off and the LED power was sequentially reduced across several daily training sessions from 10 – 0.25 mW resulting in an inverted U-shaped profile of performance as behaviour improved but stimulation powers dropped (**Figure 1 – figure supplement 3f,g**). Subsequently, animals were tested on several "high power" psychometric curve sessions (**Figure 1 – figure supplement 5f** 1.65 ± 1.06 sessions, N = 26 mice; 4 mice skipped this step)

where intermediate LED powers (250 – 50 μW) were pseudorandomly interleaved trial-by-trial (**Figure 1 – figure supplement 3h**), finally finishing Phase 1 by undergoing several "low power" psychometric curve sessions (100 – 20 μW) (**Figure 1 – figure supplement 3i – k**, **Figure 1 – figure supplement 5f** 2.12 ± 2.07 sessions, N = 26 mice; 4 mice skipped this step). Animals' response rates decreased with decreasing 1P photostimulation powers (**Figure 1 – figure supplement 3i** P(Lick) for 100 μW 0.94 ± 0.11 vs 20 μW 0.51 ± 0.29, $P = 5.96 \times 10^{-5}$ Wilcoxon signed-rank test, N = 22 mice) and their reaction times became slower (**Figure 1 – figure supplement 3j** reaction time for 100 μW 0.47 ± 0.16 s vs 20 μW 0.58 ± 0.15 s, $P = 4.61 \times 10^{-5}$ Wilcoxon signed-rank test, N = 22 mice that did this step) and increasingly variable (**Figure 1 – figure supplement 3k** reaction time s.d. for 100 μW 0.09 ± 0.04 s vs 20 μW 0.16 ± 0.08 s, $P = 5.46 \times 10^{-4}$ paired t-test, N = 22 mice that did this step), though even the lowest LED powers evoked lick rates significantly higher than catch trials (**Figure 1 – figure supplement 3i** P(Lick) for catch trials: 0.12 ± 0.08 vs go trials of 20 μW: 0.51 ± 0.29, $P = 2.31 \times 10^{-4}$; 40 μW: 0.77 ± 0.20, $P = 2.01 \times 10^{-4}$; 60 μW: 0.90 ± 0.11, $P = 2.01 \times 10^{-4}$; 80 μW: 0.94 ± 0.07, $P = 2.01 \times 10^{-4}$; 100 μW: 0.94 ± 0.11, $P = 2.01 \times 10^{-4}$; all Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons, N = 22 mice, average over 1 – 3 sessions).

At this point animals were transitioned to Phase 2 where 1P go trials (50 μW), 2P go trials (targeted to 200 and subsequently 100 neurons) and catch trials were pseudorandomly interleaved in equal proportion (**Figure 1c**,**d**). Before each training session we selected 200 neurons based on C1V1-mRuby expression in clearly expressing FOVs in superficial L2/3 barrel cortex (~130 - 230 μm below pia) and

recorded their sensitivity to photostimulation (**Figure 1 – figure supplement 4**). For each animal we selected similarly positioned FOVs across days, but did not specifically target the same neurons (which is challenging due to angular inconsistencies in FOV position across days, see *Methods*). Using these stimulus patterns for 2P go trials we trained animals on Phase 2, and subsequently some animals on Phase 3 (only 2P stim and catch trials), for several sessions (**Figure 1d** single session, **Figure 1 – figure supplement 5a,b,f** 3.40 ± 1.51 Phase 2/3 sessions 2P 200/100 neurons, N = 12 animals). During this period we also began interleaving 2P trials stimulating 100 neurons from the 200 neuron group when animals reliably detected 200 neuron stimulations (typically within a single session; 1.08 ± 0.29 sessions to d-prime > 1 on 2P 200 neuron trials, N = 12 animals) (**Figure 1 – figure supplement 5a,b** performance over time). We found that animals reliably detected 2P photostimulation of both 200 and 100 neurons consistently across time, from the first session (**Figure 1 – figure supplement 5c** first 200 neuron d-prime: 2.39 ± 0.86 vs 1, $P = 1.70 \times 10^{-4}$ paired t-test; first 100 neuron d-prime: 2.13 ± 0.93 vs 1, $P = 1.4 \times 10^{-3}$ paired t-test, N = 12 mice) to the last session (**Figure 1 – figure supplement 5c** last 200 neuron d-prime: 2.83 ± 0.62 vs 1, $P = 5.95 \times 10^{-7}$ paired t-test; last 100 neuron d-prime: 2.55 ± 0.7 vs 1, $P = 1.00 \times 10^{-5}$ paired t-test, N = 12 mice) and only showed a modest improvement over time (**Figure 1 – figure supplement 5c** 200 neuron first 2.39 ± 0.86 vs last 2.83 ± 0.62, P = 0.07 paired t-test, N = 12 mice; 100 neuron first 2.02 ± 1.28 vs last 2.86 ± 0.8, P = 0.06 Wilcoxon signed-rank test, N = 6 mice with multiple 100 neuron sessions). On their final Phase 2 session (session 2.75 ± 1.42, N = 12 animals) animals detected 2P photostimulation of 200 neurons with high response rates that were similar to 1P photostimulation (**Figure 1d** single

session, **Figure 1e** group average; $P = 1.82 \times 10^{-5}$ Friedman test; P(Lick) for 1P: 0.99 ± 0.02 vs 2P: 0.88 ± 0.10, $P = 0.19$, Bonferroni correction for multiple comparisons, $N = 12$ mice, 1 session each) and with similar reaction time mean (**Figure 1f** $P = 6.70 \times 10^{-3}$ one-way repeated measures ANOVA; 1P: 0.49 ± 0.14 s vs 2P: 0.56 ± 0.15 s, $P = 0.77$, Bonferroni correction for multiple comparisons, $N = 11$ mice, 1 session each, only mice with > 2 catch trial responses included) and standard deviation (**Figure 1g** $P = 4.99 \times 10^{-8}$ one-way repeated measures ANOVA; 1P: 0.09 ± 0.05 s vs 2P: 0.13 ± 0.02 s, $P = 0.25$, Bonferroni correction for multiple comparisons, $N = 11$ mice, 1 session each, only mice with > 2 catch trial responses included). Both 2P and 1P photostimulation evoked higher lick rates than catch trials (**Figure 1e** $P = 1.82 \times 10^{-5}$ Friedman test; P(Lick) for Catch: 0.11 ± 0.07 vs 2P: 0.88 ± 0.10, $P = 1.61 \times 10^{-2}$; vs 1P: 0.99 ± 0.02, $P = 1.04 \times 10^{-5}$, Bonferroni correction for multiple comparisons, $N = 12$ mice, 1 session each) with less variable reaction times (**Figure 1g** reaction time s.d.; $P = 4.99 \times 10^{-8}$ one-way repeated measures ANOVA; Catch: 0.28 ± 0.09 s vs 2P: 0.13 ± 0.02 s, $P = 5.99 \times 10^{-6}$; vs 1P: 0.9 ± 0.05 s, $P = 7.24 \times 10^{-8}$, Bonferroni correction for multiple comparisons, $N = 11$ mice, 1 session each, only mice with > 2 catch trial responses included), though only 1P trials showed quicker reaction times (**Figure 1f** $P = 6.70 \times 10^{-3}$ one-way repeated measures ANOVA; Catch: 0.69 ± 0.11 s vs 1P: 0.49 ± 0.14 s, $P = 5.91 \times 10^{-3}$; vs 2P: 0.56 ± 0.15 s, $P = 0.10$, Bonferroni correction for multiple comparisons, $N = 11$ mice, 1 session each, only mice with > 2 catch trial responses included). We also note that we found similar response rates in animals expressing non-somatically-restricted C1V1 (**Figure 1 – figure supplement 6a** catch subtracted P(Lick) for C1V1-Kv2.1 0.77 ± 0.11 vs C1V1 0.79 ± 0.14, $P = 0.61$ two-sample t-test, $N = 12$ C1V1-Kv2.1 and 19 C1V1 injected mice), though reaction times

were significantly slower for somatically-restricted C1V1 (**Figure 1 – figure supplement 6** C1V1-Kv2.1 0.55 ± 0.15 s vs C1V1 0.41 ± 0.08 s, $P = 2.09 \times 10^{-3}$ two-sample t-test, $N = 12$ C1V1-Kv2.1 and 19 C1V1 injected mice).

Thus we have demonstrated that two-photon optogenetic stimulation targeted to small ensembles of cortical neurons can reliably drive behaviour and provides a powerful tool for investigating the perceptual salience of different patterns of neural activity.

Very few cortical neurons are sufficient to drive behaviour

We next tested behavioural sensitivity to the activity of neural ensembles of varying size. To test this we transitioned animals to Phase 3 sessions (2P and catch trials only; **Figure 1c**: bottom) where we precisely titrated the level of activation by targeting different numbers of neurons on a trial by trial basis. We again selected 200 neurons on the basis of C1V1-mRuby expression and sub-divided this group into smaller subsets of 100, 75, 50, 25 10 and 5 neurons (**Figure 2a**). Animals then underwent 2P photostimulation psychometric curve sessions during which, in addition to stimulating the group of 200 neurons, we also pseudorandomly interleaved stimulations of the smaller subsets of neurons (**Figure 2b,c**). We leveraged our ability to simultaneously read out neural activity with calcium imaging to refine our estimate of the number of stimulated neurons according to the number of target neurons that were activated averaged across trials. To quantify activation (and suppression; see later sections) we defined thresholds for each neuron on the basis of their response distribution on correct reject catch trials (**Figure 2 – figure supplement 1b – g**, see *Methods*), where no stimulus or lick occurred, and to differentiate between target and

281 background neurons we defined 3D target zones around each 2P photostimulation
282 target co-ordinate (**Figure 2 – figure supplement 1a,h – j**, see *Methods*). This
283 resulted in numbers of activated target neurons that were 0.46 ± 0.20 times that of
284 the number of target zones (averaged across trial types) and decreased with
285 decreasing number of zones as intended (**Figure 2 – figure supplement 1i**; see
286 *Methods*).
287
288 Animals' response rates increased sigmoidally with increasing numbers of target
289 neurons activated (**Figure 2b** single animal, **Figure 2d** all sessions) and both
290 individual sessions and aggregate data across sessions were well fit by log-normal
291 sigmoid psychometric functions (**Figure 2d** grey dots/lines: individual data/fits, R^2 =
292 0.91 ± 0.09, *N* = 11 sessions, 6 animals, 1 – 2 sessions each; red line: aggregate fit,
293 R^2 = 0.72, cross-validated R^2 = 0.67 ± 0.23, *N* = 10,000 permutations, see *Methods*
294 for fit details). Using these fitted psychometric functions we estimate that animals can
295 detect the activation of a minimum of ~14 neurons at their perceptual threshold
296 (**Figure 2e** 10% point of curve: individual fits 14.2 ± 8.96 neurons, *N* = 11 sessions,
297 6 animals, 1 – 2 sessions each; aggregate fit 9.86 [95% CI: 6.09 12.8] neurons), with
298 only roughly double this number of neurons (~37) required to saturate performance
299 (**Figure 2e** 90% point of curve: individual fits 37.0 ± 11.3 neurons, *N* = 11 sessions,
300 6 animals, 1 – 2 sessions each; aggregate fit 45.1 [95% CI: 37.3 63.5 neurons). At the
301 psychometric function's 50% point (**Figure 2e** 50% point of curve: individual fits 21.8
302 ± 8.71 neurons, *N* = 11 sessions, 6 animals, 1 – 2 sessions each; aggregate fit 21.1
303 [95% CI: 17.9 23.9] neurons) this results in a very steep slope, with ~0.05 probability
304 of licking added per neuron stimulated (**Figure 2f** individual fits 0.05 ± 0.03; aggregate

fit 0.03). This is notably steeper for individual fits than the aggregate fit (0.05 ± 0.03 vs 0.03, $P = 9.77 \times 10^{-3}$ Wilcoxon signed rank test, $N = 11$ sessions, 6 animals, 1 – 2 sessions each). Mean reaction times did not vary as fewer neurons were activated (**Figure 2 – figure supplement 2a** $\beta = -0.02$, $R^2 = 0.02$, $P = 0.24$), though they did become more variable (**Figure 2 – figure supplement 2b** $\beta = -0.05$, $R^2 = 0.30$, $P = 1.41 \times 10^{-6}$). This demonstrates that animals are exquisitely sensitive to the activation of small numbers of cortical neurons and can read out surprisingly small changes in cortical activity levels.

We next addressed the question of how flexible this perceptual threshold is and how specific it is to neurons used for training during preceding sessions. After training animals to detect the activation of hundreds of barrel cortex neurons, we asked whether their ability to detect the activation of small subsets of these neurons improved across multiple subsequent days, and whether learning was specific to neurons targeted on each day. In one cohort of animals we identified and activated the same neurons reliably across multiple days (**Figure 2 – figure supplement 3a**) and measured the detection rate across sessions (**Figure 2 – figure supplement 3b**) whereas in a second cohort we moved FOV for each session and stimulated different neurons. Across all animals there was a consistent improvement in detection rate (**Figure 2 – figure supplement 3c** P(Lick) on Session 1: 0.17 ± 0.18 vs Session 2: 0.28 ± 0.25, $P = 6.29 \times 10^{-3}$ paired t-test, $N = 14$ animals testing the same 30 neurons and 5 animals testing different groups of 25 and 50 neurons) which did not differ depending on whether the same or different neurons were stimulated across sessions (**Figure 2 – figure supplement 3d** P(Lick) improvement for same: 0.12 ± 0.18 vs

different: 0.09 ± 0.18, P = 0.84 Mann Whitney U-Test, N = 14 animals testing the same 30 neurons and 5 animals testing different groups of 25 and 50 neurons).

These experiments use targeted stimulation of cortical neurons to describe the behavioural input-output function for our task, and suggest that the lower bound for detection is a small number of neurons (~14 neurons) and the psychometric function is very steep (saturating at ~37 neurons). We also demonstrate that animals' ability to detect small numbers of neurons improves with training and that this improvement is not limited to targeted neurons.

Suppression in the local network balances target activation

We took advantage of our ability to simultaneously record activity in both the targeted and untargeted 'background' neurons to investigate how local network activity influences or depends on behavioural performance. Using the same activation thresholds, suppression thresholds and target definitions described earlier (**Figure 2 – figure supplement 1**, see *Methods*) we calculated the proportion of activated and suppressed neurons on each trial averaged across trials of each type (**Figure 3 – figure supplement 1a,e**). Splitting trials by hits and misses we found that hits were associated with more activation (**Figure 3 – figure supplement 1a,b** P(activated) background on hits $4.94 \times 10^{-2} \pm 8.63 \times 10^{-3}$ vs misses $3.83 \times 10^{-2} \pm 5.04 \times 10^{-3}$ on 50 target trials, $P = 1.95 \times 10^{-3}$ Wilcoxon signed-rank test, N = 11 sessions, 6 animals, 1 – 2 sessions each) and less suppression (**Figure 3 – figure supplement 1e,f** P(suppressed) background on hits $3.53 \times 10^{-2} \pm 4.84 \times 10^{-3}$ vs misses $3.99 \times 10^{-2} \pm 5.5 \times 10^{-3}$ on 50 target trials, $P = 2.23 \times 10^{-2}$ paired t-test, N = 11 sessions, 6 animals, 1 – 2 sessions each) than misses. However, since hits are associated with

stereotyped behaviours (licking, whisking, face movements etc.), and significant movement and reward-related activity has been observed in primary sensory cortical areas (Shuler and Bear, 2006; Niell and Stryker, 2010; Musall *et al.*, 2019; Steinmetz *et al.*, 2019; Stringer *et al.*, 2019; Zatka-Haas *et al.*, 2020), we reasoned that such differences might be accounted for by the behaviours themselves irrespective of our manipulations. Indeed we found that in background neurons the level of activation recruited on hits post-photostimulation was not different from false alarms on catch trials where animals licked but no neurons were photostimulated (**Figure 3 – figure supplement 1d** P(activated) background on 50 target hit $4.92 \times 10^{-2} \pm 9.05 \times 10^{-3}$ vs catch false alarm $4.83 \times 10^{-2} \pm 1.04 \times 10^{-2}$, $P = 0.52$ paired t-test, $N = 10$ sessions, 6 animals, 1 – 2 sessions each, 1 session without any catch false alarms excluded), irrespective of how many neurons we activated (**Figure 3 – figure supplement 1a**). The amount of suppression also did not differ between hits post-photostimulation and false alarms on catch trials (**Figure 3 – figure supplement 1h** 50 neuron hit vs catch false alarm: $3.51 \times 10^{-2} \pm 5.05 \times 10^{-3}$ vs $3.37 \times 10^{-2} \pm 6.99 \times 10^{-3}$, $P = 0.27$ paired t-test, $N = 10$ sessions, 6 animals, 1 – 2 sessions each, 1 session without any catch false alarms excluded), though there was some modulation of this difference with the number of neurons activated (**Figure 3 – figure supplement 1e**). It therefore seemed possible that a significant amount of the stimulus-evoked activity that we read out in background neurons was influenced by lick-related behaviours. In line with this we found that a large fraction of neurons showed activity which was modulated by spontaneous licking (**Figure 3 – figure supplement 1i – k** 46% ± 11 of neurons lick modulated, $N = 11$ sessions, 6 animals, 1 – 2 sessions each, see *Methods*) with neurons showing both positive correlation (**Figure 3 – figure supplement 1j,l,m** 2.83

x 10^{-2} ± 1.81 x 10^{-2} lick correlation for all positively lick modulated neurons, *N* = 9547 neurons) and negative correlation (**Figure 3 – figure supplement 1j,l,m** -2.52 x 10^{-2} ± 1.51 x 10^{-2} lick correlation for all negatively modulated neurons, *N* = 4365 neurons).

Unfortunately, the temporal resolution of calcium imaging does allow us to tease apart the direction of causality in our data**,** i.e. whether this activity causes, or is caused by motor output. However given the literature demonstrating behavioural output-related activity in sensory cortices (Musall et al., 2019; Steinmetz et al., 2019; Stringer et al., 2019), and the fact that manipulating this activity has no effect on behavioural choices (Zatka-Haas et al., 2020), we were concerned that a significant amount of network activity might result from movement rather than causing it. This would be problematic for our interpretation since the amount of lick contamination will vary by trial type (no. neurons activated) in a manner that correlates with the variable under study (due to the increased P(Lick) with no. neurons activated **Figure 2d**). To take account of this we devised a hit/miss matching procedure which removes the variance in hit:miss ratio across trial types by ensuring that all trial types have a 50:50 ratio of hits:misses (**Figure 3 – figure supplement 1o**, see *Methods*). This is achieved for trials of a given type (i.e. a low P(Lick) trial type: 10 activated neurons) by matching the number of trials of the minority response type (i.e. Hits) with random resamples, of the same number, of majority response-type trials (i.e. Misses) and averaging network response metrics across resamples. Following this procedure should ensure that all trial types have the same proportion of data contaminated by lick responses and any variation in network response across trial types remaining should be due to the variation in number of target neurons activated.

Using this procedure, we investigated how the network response varies as a function of stimulated ensemble size beyond its stereotyped modulation by the behavioural response. Taking the network as a whole (including target neurons) we found that photostimulation causes both activation (**Figure 3a** right inset; P(activated) all neurons on photostimulus $5.91 \times 10^{-2} \pm 1.20 \times 10^{-2}$ vs catch $4.82 \times 10^{-2} \pm 6.65 \times 10^{-3}$ trials averaged across all trial types, $P = 1.17 \times 10^{-3}$ paired t-test, $N = 10$ sessions, 6 mice, 1 – 2 sessions each) and suppression (**Figure 3b** right inset; P(suppressed) all neurons on photostimulus $4.58 \times 10^{-2} \pm 4.90 \times 10^{-3}$ vs catch $3.97 \times 10^{-2} \pm 4.80 \times 10^{-3}$ trials averaged across all trial types, $P = 1.41 \times 10^{-4}$ paired t-test, $N = 10$ sessions, 6 mice, 1 – 2 sessions each) and that both scale with the number of neurons activated (**Figure 3a,b** P(activated) all neurons: $\beta = 7.54 \times 10^{-3}$, $R^2 = 0.24$, $P = 2.36 \times 10^{-5}$; P(suppressed) all neurons: $\beta = 3.00 \times 10^{-3}$, $R^2 = 0.25$, $P = 1.54 \times 10^{-5}$, $N = 10$ sessions, 6 mice, 1 – 2 sessions each). When we analysed only background neurons (i.e. excluding targets from the calculation) we found that while photostimulation does cause both activation and suppression in the background network (**Figure 3d,e** right insets; P(activated) network on photostimulus $4.40 \times 10^{-2} \pm 6.89 \times 10^{-3}$ vs catch $4.02 \times 10^{-2} \pm 5.60 \times 10^{-3}$ trials, $P = 2.31 \times 10^{-3}$ paired t-test; P(suppressed) background on photostimulation $3.81 \times 10^{-2} \pm 4.60 \times 10^{-3}$ vs catch $3.27 \times 10^{-2} \pm 4.60 \times 10^{-3}$ trials, $P = 1.06 \times 10^{-4}$ paired t-test, all averaged across trial types, $N = 10$ sessions, 6 mice, 1 – 2 sessions each), only background network suppression scales with the number of activated target neurons (**Figure 3e** $\beta = 2.95 \times 10^{-3}$, $R^2 = 0.31$, $P = 7.79 \times 10^{-7}$) whereas activation does not (**Figure 3d** $\beta = 1.70 \times 10^{-3}$, $R^2 = 0.05$, $P = 0.08$). Moreover, activation and suppression have distinct spatial profiles, with significant suppression

occurring over a much broader area (**Figure 3 – figure supplement 3**). This suggests that the network reacts to suppress the spread of activation triggered by our photostimulation in a graded manner which tracks the activation strength, whereas network activation changes to a smaller extent. Indeed, across all neurons in the population we see that there is a consistent balance of activation and suppression across all target activation levels (**Figure 3c** P(Activated)/P(Suppressed) across all neurons $\beta = 7.23 \times 10^{-2}$, $R^2 = 0.05$, $P = 0.06$) that remains similar to the rates observed spontaneously during catch trials (**Figure 3c** right inset; P(Activated)/P(Suppressed) on photostimulation 1.30 ± 0.28 vs catch 1.24 ± 0.25 trials across all neurons, $P = 0.25$ paired t-test, $N = 10$ sessions, 6 mice, 1 – 2 sessions each).

These results suggest that activating target neurons produces suppression in the surrounding network which maintains homeostasis of background activity.

Behaviour tracks target neuron activity despite constant, matched suppression in the local network

As it has been suggested that both excitation and inhibition of cortical neurons can drive behaviour (Doron, von Heimendahl, Schlattmann, Houweling, & Brecht, 2014), and since target activation and network suppression correlate in our dataset (**Figure 3e**), we finally asked which of the factors that we have analysed best correlates with animals' behavioural responses. Using the hit:miss matched data described above we modelled behavioural response rates as a function of target activation, background activation and background suppression (**Figure 4a – c**), cross-validating across training and test datasets to assess the goodness and generalisability of the fits (see *Methods*). The strongest and most generalisable predictor of behavioural

responses was target neuron activation, which had significantly positive R^2 across both training and test sets during cross-validation (**Figure 4a,d** train R^2: 0.71 ± 0.06, P = 0 permutation test vs 0; test R^2: 0.66 ± 0.23, $P = 1.68 \times 10^{-2}$ permutation test vs 0, N = 10,000 train:test splits). Network suppression had mild predictive power on training data, but none on testing data (**Figure 4c,d** train R^2: 0.16 ± 0.06, $P = 2.50 \times 10^{-3}$ permutation test vs 0; test R^2: 0.01 ± 0.39, P = 0.39 permutation test vs 0, N = 10,000 train:test splits) which might be explained by its correlation with target activation (**Figure 3e**). Network activation was a poor predictor of both training and testing data, suggesting that it had little influence on behavioural performance (**Figure 4b,d** train R^2: -0.05 ± 0.05, P = 0.11 permutation test vs 0; test R^2: -0.21 ± 0.44, P = 0.22 permutation test vs 0, N = 10,000 train:test splits). Thus within the cortical region that we can observe and manipulate we have tested three of the main pathways by which activity in cortex could influence downstream circuitry to ultimately drive behaviour: (1) output of directly activated target neurons; (2) output of background neurons synaptically activated by target neurons; (3) suppression of the output of background neurons through disynaptic inhibition by interneurons activated by target neurons (**Figure 4e**). Our results indicate that the most robust effect is the number of activated target neurons (1), with only a minor impact of indirectly modulated neurons in the local network.

Discussion

By precisely titrating the number of activated neurons to be detected, we have demonstrated that the psychometric function for detecting cortical activity is sensitive, with only tens of neurons required to drive behaviour, and it is steep, with only an approximate doubling in number sufficient to drive asymptotic behavioural performance. Simultaneous imaging of the surrounding network has allowed us to show that, despite this exquisite behavioural sensitivity, the dominant network response matches the activation of targeted neurons with local suppression, flattening the network input-output function to maintain the level of network activation within the spontaneous range. These results support the sparse coding hypothesis (Barlow, 1972; Barth & Poulet, 2012; Kanerva, 1993; Olshausen & Field, 1996), demonstrate that the local network operates in an inhibition-stabilized regime (Denève & Machens, 2016; Murphy & Miller, 2009; Ozeki, Finn, Schaffer, Miller, & Ferster, 2009; Pehlevan & Sompolinsky, 2014; Sanzeni et al., 2020; Tsodyks et al., 1997; van Vreeswijk & Sompolinsky, 1996; Wolf et al., 2014) and suggest a high storage capacity for recurrent networks (Hopfield, 1982; Lefort *et al.*, 2009; Peron *et al.*, 2020). This combination of features likely maximizes perceptual sensitivity while minimizing erroneous detection of background activity, thus avoiding hallucinations (Carbon, 2014; Cassidy et al., 2018; Corlett et al., 2019; Friston, 2005), runaway excitation (Rose & Blakemore, 1974; Treiman, 2001; Ziburkus, Cressman, Barreto, & Schiff, 2006) and reducing cortical energy requirements (Schölvinck et al., 2008). We also show that the lower bound of detectable activity is not fixed, but can improve with training in a way that can generalise to other neurons in the surrounding network. This suggests that perceptual learning can reroute cortical resources to meet even

the most stringent task demands and generalise to other potentially relevant neurons, increasing cognitive flexibility. This also demonstrates that the brain's ability to "learn to learn" (Behrens et al., 2018; Harlow, 1949) can extend even to arbitrary activity patterns.

Activation of only a small number of neurons is required to reach perceptual threshold

We leveraged our all-optical system to precisely target different numbers of neurons for two-photon optogenetic activation in order to define the minimum number that is sufficient to drive behaviour. The perceptual threshold we measured is remarkably low: mice can detect the activation of only ~14 cortical neurons. This number is substantially lower than the threshold estimated for one-photon activation of layer 2/3 neurons in rodent barrel cortex (~60; Huber *et al.*, 2008). This might be explained by the fact that we tailor our optogenetic stimulus for activation of individual neurons, whereas one-photon stimulation diffusely activates an entire population of neurons (many of which will receive subthreshold levels of photocurrent). Moreover, our stimuli drive double the number of action potentials (~10 in each neuron) which Huber *et al.* show increases detectability, though we note that our minimum threshold of ~140 action potentials (10 action potentials in 14 neurons) is still lower than the ~300 that they report. Furthermore, we produce comparatively more clustered activation of stimulated neurons in our ensembles (confined to a ~500 x 500 x 100 μm volume) compared to the more dispersed neuron locations used in Huber *et al.* (potentially across the whole of S1), which may result in different recruitment of both local and downstream targets due to changes in intra- and inter-laminar connection probability

519 with distance (Holmgren, Harkany, Svennenfors, & Zilberter, 2003; Lefort et al., 2009;
520 Perin, Berger, & Markram, 2011; Thomson & Lamy, 2007; Yoshimura, Dantzker, &
521 Callaway, 2005). Finally, the Huber *et al.*, 2008 study relied on post-hoc histological
522 estimates which, as noted by the authors, may have a significant margin of error.
523
524 On the other hand, our estimate of the perceptual threshold is an order of magnitude
525 higher than single-cell patch-clamp experiments demonstrating that strong activation
526 of single neurons can in some cases lead to a behavioural report (Doron et al., 2014;
527 Houweling & Brecht, 2008; Tanke, Borst, & Houweling, 2018). Several key differences
528 could account for this discrepancy. First, these studies mostly stimulated L5 neurons,
529 which serve as the output neurons of the cortical circuit and therefore may drive
530 behaviour more reliably compared to the L2/3 ensembles we target. Indeed, such a
531 dichotomy is confirmed by a recent study comparing the behavioural influence of
532 functionally defined ensembles in L2/3 with those in L5 (Marshel et al., 2019).
533 Secondly, these studies report significant variability in the ability of single neurons to
534 drive behaviour, with many neurons having no effect, and with the most potent
535 behavioural effects being limited mainly to fast-spiking putative interneurons (Doron
536 et al., 2014). Additional factors which may play a role include differences in level of
537 stimulation of individual neurons, differences in the training protocol, or species
538 differences between mice and rats.
539
540 Interestingly, the reaction times to optogenetic stimuli that we report (0.4 – 0.7 s) are
541 comparable to, though slightly slower than, those reported for detection of whisker
542 stimuli in mice (0.3 – 0.4 s; J. L. Chen, Carta, Soldado-Magraner, Schneider, &

Helmchen, 2013; Hires, Gutnisky, Yu, O'Connor, & Svoboda, 2015; O'Connor et al., 2010; Sachidhanandam, Sreenivasan, Kyriakatos, Kremer, & Petersen, 2013) suggesting that they may be processed differently. This could be because sensory stimuli more robustly recruit many neurons distributed over several parallel thalamo-cortical pathways which provide input to multiple cortical layers (Feldmeyer et al., 2013; Petersen, 2007), including direct projections from thalamus to cortical output L5 (Constantinople & Bruno, 2013), whereas our optogenetic stimuli will only activate small numbers of neurons in L2/3. Moreover, sensory stimuli will likely drive patterns of activity that respect cortical wiring for transmission of sensory information and so may propagate more potently downstream. Future work comparing reaction times to targeted activation of sensory-evoked or random ensembles of neurons in the same animal will yield insight into any distinction between naturalistic and artificial neural activity in driving behaviour. Our results also reveal reduced reaction time for non-somatically-restricted C1V1 with little difference in response rate. This difference in reaction time may be due to increased off-target activation via processes traversing the photostimulation volume as well as larger photocurrent in target neurons through activation of opsin in neuronal compartments in addition to the soma (dendrites/axon). It is possible that this does not result in an increased response rate because animals' performance is saturated at its upper bound (allowing for a lapse rate that is independent of the salience of stimuli).

What are the functional implications of the perceptual threshold we have defined? The fact that in our study (see also Daie, Svoboda, & Druckmann, 2019; Gill et al., 2020; Marshel et al., 2019) the lower perceptual bound is well above a single neuron

suggests that perceptual thresholds are tuned to minimize conscious perception of the spontaneous neural activity which often co-exists alongside stimulus activity (Bernander, Douglas, Martin, & Koch, 1991; Destexhe & Pare, 1999; Destexhe, Rudolph, & Paré, 2003; London, Roth, Beeren, Häusser, & Latham, 2010; Musall et al., 2019; Shadlen & Newsome, 1994; Stringer et al., 2019; Tolhurt, Movshon, & Dean, 1983; Waters & Helmchen, 2006). If the perceptual apparatus is sensitive to single-neuron activation this may lead to erroneous detection of background cortical activity. While the stimuli in our study do not explicitly mimic sensory evoked activity, such false positives in our behavioural paradigm may be related to the hypothesized role of false sensory percepts in generating hallucinations, which impair normal cognitive function and are associated with an array of pathological conditions (Chaudhury, 2010; Kumar, Soren, & Chaudhury, 2009; Llorca, Pereira, Jardri, Brousse, & Misdrahi, 2016). The threshold of ~14 neurons could therefore be important for avoiding perceptual false positives caused by spontaneous background network activity.

The relatively low perceptual threshold may also have significant computational advantages. A large body of theoretical work has suggested that the brain may use a sparse coding scheme to represent information (Barlow, 1972; Kanerva, 1993; Olshausen and Field, 1996). Our demonstration that the perceptual threshold (~14 neurons) is much lower than the dimensionality of barrel cortex (~400,000 neurons in barrel cortex; Hooks *et al.*, 2011; Meyer *et al.*, 2013; ~2000 neurons in superficial layers of a barrel; Lefort *et al.*, 2009), but also significantly higher than a single neuron, is consistent with computational theories proposing that individual items are

represented sparsely compared to the dimensionality of the space, but also that they are not represented by single elements of that space (Baum, Moody, & Wilczek, 1988; Kanerva, 1993; Olshausen & Field, 2004; Palm, 1980). Our work is also consistent with work suggesting that barrel cortex can use ensembles on this scale to robustly encode sensory information (Hires et al., 2015; Mayrhofer, Haiss, Helmchen, & Weber, 2015; Panzeri, Ince, Diamond, & Kayser, 2014; Stüttgen & Schwarz, 2008). Such sensitivity is beneficial as it suggests that recurrent networks like cortical L2/3 have a high storage capacity (Hopfield, 1982; Lefort *et al.*, 2009; Ko *et al.*, 2011; Harris and Mrsic-Flogel, 2013; Cossell *et al.*, 2015; Peron *et al.*, 2020) allowing the brain to represent many patterns independently (Amit, Gutfreund and Sompolinsky, 1985a; Amit, Gutfreund and Sompolinsky, 1985b; McEliece *et al.*, 1987; Brunel, 2016; Folli, Leonetti and Ruocco, 2017).

The network input-output function for perception is steep

By recording the response of the local network while carefully titrating the number of targeted neurons we have defined the network input-output function for perception in our task. This function is sigmoidal and remarkably steep, saturating at only ~37 neurons. These results echo similarly steep perceptual input-output functions found in other systems (Gill et al., 2020; Marshel et al., 2019) but are much steeper than estimated in barrel cortex for one-photon optogenetic stimulation (Huber et al. 2008). Again, this discrepancy may be due to a range of factors associated with one-photon photostimulation, from the spatially diffuse (and largely subthreshold) nature of the activation to differences in network cooperativity associated with our more clustered stimulation patterns. Indeed, the highly synergistic activation of a local network of

densely interconnected excitatory neurons (Cossell et al., 2015; Douglas, Koch, Mahowald, Martin, & Suarez, 1995; Ko et al., 2011), rapidly followed by suppression in the local network (London et al. 2010; Kwan & Dan, 2012; Chettih & Harvey, 2019) likely mediated by disynaptic inhibition (Jouhanneau, Kremkow, & Poulet, 2018; Mateo et al., 2011; Silberberg & Markram, 2007), may be the basis for the steep and saturating input-output function we have observed.

This steep input-output function may have significant functional consequences. While allowing rejection of noise due to spontaneous activity (see the previous section), it also enables perceptual detection of relevant activity with high sensitivity and efficiency, and yet avoids further unnecessary engagement of the network with additional stimulation. This may represent a circuit mechanism for optimizing the canonical trade-off in a sensory system subject to noise (Bialek, 2012) between minimizing false positives (a response when there is no signal: a “false alarm”) and minimizing false negatives (missing a signal when there is one present). The steepness of the input-output function and the low number of neurons at saturation are also consistent with optimal energy efficiency (Attwell & Laughlin, 2001; Lennie, 2003). Indeed, our results offer cellular-resolution support for the proposal that the energy associated with conscious perception is surprisingly low (Schölvinck et al., 2008), since our data predict that the number of additional neurons required to allow a subconsciously processed sensory stimulus to be consciously perceived will be low.

While our behavioural paradigm relies on stimulation of artificially defined ensembles of neurons, we maintain that our results add general insight as to how neural activity can underlie flexible behaviour in barrel cortex since: (1) bulk optogenetic activation of pyramidal neurons in sensory areas can readily replace trained sensory stimuli with minimal behavioural impact (Ceballo, Piwkowska, et al., 2019; O'Connor et al., 2013; Sachidhanandam et al., 2013), suggesting that the activity evoked is not so alien as to confuse behavioural processing; (2) the order of magnitude of the numbers that we report corresponds closely with the estimated number of barrel cortex neurons required to decode tactile stimuli of various types (Hires et al., 2015; Mayrhofer et al., 2015; Stüttgen & Schwarz, 2008) suggesting that there is limit to this system's sensitivity that can be found through both observation and causal manipulation; (3) irrespective of the exact numbers, the steepness of the psychometric function suggests a fine distinction between whether neural activity is perceptible or not which is indicative of a highly sensitive yet specific sensory system, as has also been shown for optogenetic stimuli mimicking sensory (visual) percepts (Marshel et al., 2019); (4) the unique ability of two-photon optogenetics to specifically target the same, or different, ensembles of neurons throughout learning has explicitly demonstrated the flexibility of this perceptual threshold and how this flexibility can generalise; (5) our ability to image background neurons in the surrounding network has added a further layer of understanding to seminal papers in the field (Houweling & Brecht, 2008; Huber et al., 2008) and demonstrates that cortical networks largely balance increasing levels of activation with matched suppression.

Nevertheless it is important to keep in mind caveats inherent to current all-optical approaches that may influence these results, such as limitations in spike readout with calcium indicators (T.-W. Chen et al., 2013; Pachitariu, Stringer, & Harris, 2018), the photostimulation efficiency of two-photon optogenetic activation (reported here and in Mardinly et al., 2018; Marshel et al., 2019; Shemesh et al., 2017) and the fact that calcium imaging subsamples the full extent of neural activity involved in complex behaviours. It is also likely that our results will be influenced by the exact stimulation parameters, such as stimulus duration, strength and timing, as has been noted for the detectability of direct cortical activation in the past (Doron et al., 2014; Gill et al., 2020; Histed & Maunsell, 2014; Huber et al., 2008). Additionally, as with all studies using trained non-naturalistic behaviour, it is worth considering how training duration might influence our results. We note that animals learn our basic task very quickly and that they continue to quickly generalise learning to new, harder stimuli over time as is often observed in trained sensory paradigms (Andermann, Kerlin, & Reid, 2010; Gerdjikov, Bergner, Stüttgen, Waiblinger, & Schwarz, 2010; O'Connor et al., 2010). This quick learning means that their performance is stable and effectively saturated by the time we test their perceptual sensitivity, in line with the way in which sensitivity to sensory stimuli is tested in many systems (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Busse et al., 2011; Carandini & Churchland, 2013; Morita, Kang, Wolfe, Jadhav, & Feldman, 2011). Therefore, our results are interpretable within this standard framework of testing perceptual sensitivity in animals that have learned a non-naturalistic task close to saturation (though see next section), be it contingent on sensory or artificial stimulation. Moreover, as we explore in the next section, we demonstrate that this threshold can change with training, suggesting that the more

pertinent feature to consider in relation to perception more generally may be the steepness of animals' psychometric curves. Indeed, our results using artificial ensembles complement the psychometric functions reported in recent studies driving behaviour by optogenetically mimicking sensory ensembles (Carrillo-Reid et al., 2019; Marshel et al., 2019). Finally, an additional factor influencing our experiments is that bulk 1P optogenetic activation over long training periods may change activity and connectivity patterns, as has been suggested for repeated exposure to 2P optogenetic stimulation of the same neurons over time in the absence of behaviour (Carrillo-Reid et al., 2016) and for repeated exposure to the same sensory stimuli over behavioural training (J. L. Chen et al., 2015; Khan et al., 2018; Peron, Freeman, Iyer, Guo, & Svoboda, 2015; Poort et al., 2015; Wiest, Thomson, Pantoja, & Nicolelis, 2020). Future work recording neural responses in the population before, during and after 1P/2P behavioural training will yield important insight into this process and it will be crucial to compare the changes observed between animals trained on optogenetic stimuli and sensory stimuli.

The perceptual threshold is plastic and can generalise

We have used our ability to specifically target the same (or different) neurons across multiple days to show that the perceptual threshold is not fixed and depends on learning in a neuron-agnostic manner. This suggests that the perceptual apparatus can flexibly adapt in order to adjust the trade-offs between different kinds of errors while maximizing sensitivity (e.g. minimizing false positives vs false negatives). It also underscores the brain's ability to reroute its resources (J. L. Chen et al., 2015; Hong, Lacefield, Rodgers, & Bruno, 2018; Huber et al., 2012; Kawai et al., 2015; Law & Gold,

2008; Ölveczky, Otchy, Goldberg, Aronov, & Fee, 2011) to adaptively meet task demands with ever increasing sensitivity and accuracy (Carandini & Churchland, 2013; Fahle, 2002; Gilbert, Sigman, & Crist, 2001; Sasaki, Nanez, & Watanabe, 2009). Previous studies addressing this question in the context of sensory tasks have suggested that such learning is associated with changes in the representation of sensory stimuli in primary sensory areas (J. L. Chen et al., 2015; Khan et al., 2018; Peron et al., 2015; Poort et al., 2015; Wiest et al., 2020). However in our case animals learn despite the stimulus (direct stimulation) being held constant in sensory cortex. We therefore hypothesise that, in our task, such learning-related changes likely occur in downstream regions like S2 (J. L. Chen et al., 2015; Kwon, Yang, Minamisawa, & O'Connor, 2016), motor cortex (J. L. Chen et al., 2015; Huber et al., 2012) or striatum (Sippy, Lapray, Crochet, & Petersen, 2015; Xiong, Znamenskiy, & Zador, 2015).

The fact that the learning we observe during the 2P training phase generalizes to neurons that are not stimulated during this period suggests that animals can "learn to learn" (Harlow, 1949) within the context of detecting arbitrary cortical activity patterns, similar to that observed in tasks relying on more naturalistic neural processing (Fahle, 2002; Rudebeck & Murray, 2011; Tse et al., 2007; Walton, Behrens, Buckley, Rudebeck, & Rushworth, 2010). Such generalization of knowledge acquired from one learning epoch to another is a hallmark of the type of powerful statistical learning systems that could underlie some of the brain's most complex, flexible behaviours (Behrens et al., 2018; Eichenbaum & Cohen, 2014; Fahle, 2002; Gustafson & Daw, 2011; Stachenfeld, Botvinick, & Gershman, 2018; Tolman, 1948; Tolman, Ritchie, & Kalish, 1946; Whittington, Muller, Barry, & Behrens, 2018;

Whittington et al., 2019). Our results, and the experimental paradigm that we present, could provide a useful framework to further investigate how learning and credit are assigned to ensembles of neurons in an appropriate yet generalizable way.

The relationship between the perceptual threshold that we measure and how much learning can generalise also merits further investigation. It would be interesting to investigate how nearby stimulated neurons have to be to previously trained neurons, either physically or in terms of tuning similarity, for learning to generalise to them. Indeed another all-optical study hints that generalisation is limited to neurons that share stimulus tuning congruent with ensembles of neurons that have been previously trained (Marshel et al., 2019) implying that generalisation is not a universal property. Moreover, it would also be interesting to see whether the strength of generalisation scales with the amount of uncertainty in the preceding trained stimulus set. One could imagine that learning would be more likely to generalise if different neurons were activated on each training day (as in our experiments), or even each trial, than if just one pattern was trained for the same duration. Such volatility in the learning environment is indeed thought to change learning dynamics (Behrens, Woolrich, Walton, & Rushworth, 2007; Massi, Donahue, & Lee, 2018; Mcguire, Nassar, Gold, & Kable, 2014) and by extension may influence the level of generalisation at the neural level, as has recently been suggested for hippocampal representations (Plitt & Giocomo, 2019; Sanders, Wilson, & Gershman, 2020).

How might such generalisability of learning arise? It could in part result from increased connectivity between opsin-expressing neurons through plasticity induced

during their synchronous activation during 1P training phases, equivalent to how artificial subnetworks might be generated by 2P all-optical methods (Carrillo-Reid et al., 2016; Zhang, Russell, Packer, Gauld, & Häusser, 2018, though see also Alejandre-García et al., 2020 for alternative non-Hebbian mechanisms). In this case, subsequent 2P photostimulation of opsin-expressing neurons on a given day might preferentially recurrently excite other non-targeted opsin-expressing neurons on that day to a greater extent than non-opsin-expressing neurons, as is thought to happen with 2P optogenetic recall of artificially generated subnetworks (Carrillo-Reid et al., 2016). This could cause them to become active and "bound into" the learning process on that day, despite not being directly targeted, and allow them to better drive behaviours when targeted on subsequent days. This would be an intriguing mechanism and, while such changes in connectivity in our task would be "artificial", other work suggests that similar changes might underlie generalisation in more natural sensory guided tasks. Neurons sharing functional tuning to sensory stimuli tend to form recurrently connected subnetworks (Carrillo-Reid et al., 2019; Chettih & Harvey, 2019; Cossell et al., 2015; Jennings et al., 2019; Ko et al., 2013; Marshel et al., 2019; Peron et al., 2020; Russell et al., 2019; Znamenskiy et al., 2018), which result in non-targeted members being recruited when a subset are targeted for photostimulation (Carrillo-Reid et al., 2019; Jennings et al., 2019; Marshel et al., 2019; Russell et al., 2019), and learning can preferentially generalise across neurons within such subnetworks in sensory-guided tasks (Marshel et al., 2019). Thus while the subnetworks that might be generated through our 1P training may be artificial, the process of learning generalisation that we observe may also occur in more naturalistic sensory-driven tasks. Indeed, the fact that this mechanism can extend beyond

naturalistic stimuli to aid in detection of arbitrary stimulus patterns speaks to how pivotal it may be in helping the brain generate flexible behaviour.

The combination of sensitivity and flexibility that we report also raises the question of whether animals could be rapidly trained to detect the activity of small numbers of neurons *de novo,* without prior conditioning. This may be difficult to demonstrate in a realistic experimental timeframe given that animals have a tendency to adopt easy but sub-optimal strategies, such as timing licks to coincide with the mean of the trial-time distribution, when they are faced both with non-naturalistic task design and stimuli that are hard to detect/discriminate. Once these strategies are adopted, such local optima tend to be very hard to train away and are thus better avoided in the first place. Indeed, studies testing perceptual sensitivity to or discrimination of sensory stimuli overwhelmingly begin with easier stimulus types to habituate the animal to the novel task at hand and learning continues as more difficult stimuli are introduced (Abraham et al., 2004; Andermann et al., 2010; Busse et al., 2011; Carandini & Churchland, 2013; Gerdjikov et al., 2010; Histed, Carvalho, & Maunsell, 2012; Lee et al., 2012; Morita et al., 2011; O'Connor et al., 2010). These features of sensory-evoked behavioural performance are analagous to how animals in our task constantly adapt over time to reductions in stimulus strength, even down to the lowest stimulus levels tested. Furthermore, all previous studies using all-optical techniques to influence behaviour with cellular resolution optogenetics have incorporated some kind of conditioning phase using either sensory or optogenetic stimuli of progressively lower strength (Carrillo-Reid et al., 2019; Gill et al., 2020; Jennings et

al., 2019; Marshel et al., 2019; Russell et al., 2019) again implying that this may be necessary when probing the limits of perception.

Perception is sensitive despite matched network suppression

The matched suppression that we observed in the local L2/3 network is in accordance with the general net inhibitory effect of pyramidal neuron stimulation observed *in vivo* (Chettih & Harvey, 2019; Kwan & Dan, 2012; Mateo et al., 2011; Russell et al., 2019) and in detailed network models of cortex (Cai et al., 2020). This supports the idea that such networks operate in an inhibition-stabilized regime where one role of inhibition is to control strong recurrent excitation (Denève & Machens, 2016; Murphy & Miller, 2009; Ozeki et al., 2009; Pehlevan & Sompolinsky, 2014; Sanzeni et al., 2020; Tsodyks et al., 1997; van Vreeswijk & Sompolinsky, 1996; Wolf et al., 2014), though since our perturbations are not targeted to inhibitory neurons with specific tuning we cannot assess how functionally specific this architecture might be (Sadeh & Clopath, 2020). However, these results also seemingly challenge recent work demonstrating that activation of co-tuned ensembles in V1 predominantly activates other similarly tuned neurons in the surrounding network (Carrillo-Reid et al., 2019; Marshel et al., 2019) and that ablation of some neurons within functional sub-groups reduces activity in the spared neurons (Peron et al., 2020). This discrepancy is likely explained by the fact that recurrent excitation is known to increase with tuning similarity such that neurons sharing functional tuning tend to recurrently excite each other (Cossell et al., 2015; Ko et al., 2011) whereas inhibition is generally less tuned and structured (Kerlin *et al.*, 2010; Bock *et al.*, 2011; Fino and Yuste, 2011; Hofer *et al.*, 2011; Packer and Yuste, 2011; Scholl *et al.*, 2015,

though see Ye *et al.*, 2015; Znamenskiy *et al.*, 2018). Indeed, Marshel *et al.*, 2019 specifically use a V1 network model relying on recurrent excitation between co-tuned neurons and strong general inhibition, which keeps activity in check, to explain the low threshold and steepness of their psychometric functions. The subset of neurons we targeted, which may not share functional tuning, are unlikely to benefit from such preferential recurrent connectivity but they will likely recruit general inhibition (though as mentioned above some form of enhanced connectivity or intrinsic excitability may have been induced during early 1P training). Therefore, the matched suppression and steep psychometric functions that we observe are consistent with this model.

Since it has not been possible up until very recently to assess the impact of such titrated activation of cortical neurons on the local network during behaviour (Doron et al., 2014; Histed & Maunsell, 2014; Houweling & Brecht, 2008; Huber et al., 2008; Tanke et al., 2018, though see recent work Gill et al., 2020; Marshel et al., 2019), recent theoretical work inspired by previous behavioural results has explored how simulated neural networks can detect the activation of single neurons (Bernardi et al., 2020; Bernardi & Lindner, 2017, 2019). These studies make three key predictions: (1) the pool of readout neurons must be biased in favour of connecting with the stimulated neuron (Bernardi & Lindner, 2017), (2) the readout network must include local recurrent inhibition to mitigate noise-inducing neuronal cross-correlations (Bernardi & Lindner, 2019) and (3) inhibition must lag excitation in the readout network (Bernardi et al., 2020). Since our paradigm allows us to simultaneously monitor the local network response during behavioural detection of similarly sparse activity, we can assess the validity of such predictions *in vivo*. The strong suppression recruited

by our stimulation suggests that powerful inhibition is at work in the network and therefore supports prediction (2). The activity we induce drives behaviour despite this strong local suppression, suggesting that excitation might be transmitted to downstream circuits responsible for driving behaviour before inhibition has a chance to quell it locally. This supports prediction (3). Our results concerning prediction (1) are more mixed. Bernardi and Lindner, 2017 suggest that the predicted bias could arise due to Hebbian plasticity between stimulated and readout neurons during the initial microstimulation phase of training. They take as evidence for this the fact that naïve animals cannot detect single neurons, something which both we and other similar studies also see (Carrillo-Reid et al., 2019; Gill et al., 2020; Histed & Maunsell, 2014; Huber et al., 2008; Marshel et al., 2019). The fact that animals generally require initial one-photon priming before being able to detect targeted two-photon stimuli therefore lends some support to prediction (1). However, somewhat contradictory to this prediction is our finding that the amount by which animals improve in their detection of threshold stimuli across sessions is similar irrespective of whether the same or different neurons were stimulated. The bias in connectivity that supposedly develops between target neurons and readout neurons should not extend to other neurons that are not targeted on that day. The fact that we observe such a transfer, manifested in a similar learning rate across different neurons targeted across days, suggests that this bias may be more general than hypothesized above and may apply across most neurons contained within the area where learning has taken place, potentially via recurrent lateral connectivity (either existing or induced during 1P training).

Outlook

The combination of techniques that we have deployed provide a powerful experimental framework that can be used to test how more nuanced features of cellular identity and specific patterns of cortical activity influence perception. This will bring us closer to the goal of testing precisely which of the candidate features of the neural code underlie the considerable flexibility and processing power of the brain (Jazayeri & Afraz, 2017; Panzeri, Harvey, Piasini, Latham, & Fellin, 2017).

Materials and Methods

All experimental procedures were carried out under Project Licence 70/14018 (PCC4A4ECE) issued by the UK Home Office in accordance with the UK Animals (Scientific Procedures) Act (1986) and were also subject to local ethical review. All surgical procedures were carried out under isoflurane anaesthesia (5% for induction, 1.5% for maintenance), and every effort was made to minimize suffering.

Animal preparation

4 – 6 week old wild-type (C57/BL6) and transgenic GCaMP6s mice (Emx1-Cre;CaMKIIa-tTA;Ai94) of both sexes were used. A calibrated injection pipette (15 μm inner diameter) bevelled to a sharp point was mounted on an oil-filled hydraulic injection system (Harvard apparatus) and front-loaded with virus (either a 1:10 mixture of AAV1-Syn-GCaMP6s-WPRE-SV40 and AAVdj-CaMKIIa-C1V1(E162T)-TS-P2A-mCherry-WPRE or a 1:8 mixture of AAV1-Syn-GCaMP6s-WPRE-SV40 and either AAV2/9-CaMKII-C1V1(t/t)-mScarlett-Kv2.1 or AAV2/9-CaMKII-C1V1(t/t)-mRuby2-Kv2.1). AAV2/9-CaMKII-C1V1(t/t)-mScarlett-Kv2.1 and AAV2/9-CaMKII-C1V1(t/t)-mRuby2-Kv2.1 virus was diluted in virus buffer solution (20 mM Tris, pH 8.0, 140 mM NaCl, 0.001% Pluronic F-68) 10-fold relative to stock concentration (~6.9×10^{14} gc/ml). These constructs were as in Chettih and Harvey, 2019. One of the latter two somatically restricted (Kv2.1) opsins was used for all experiments except those targeting the same neurons across days (**Figure 2 – figure supplement 3**) where non-restricted opsin was used. Of the 22 mice initially trained on 1P stimulation, 4 were opsin-injected GCaMP6s transgenics and 18 were opsin/indicator-injected WT mice. Of these, 6 mice were used for 2P psychometric curve experiments (**Figure 2**

908 **– 4**), 4 of these were opsin-injected GCaMP6s transgenics and 2 were
909 opsin/indicator-injected WT mice (see *Behavioural training* below for details of mice
910 used for each training phase). Mice were given a peri-operative sub-cutaneous
911 injection of 0.3 mg/mL buprenorphine hydrochloride (Vetergesic). They were then
912 anaesthetized with isoflurane (5% for induction, 1.5% for maintenance) and the scalp
913 above the dorsal surface of the skull was removed. A metal headplate with a 7 mm
914 diameter circular imaging well was fixed to the skull over right S1 (2 mm posterior and
915 3.5 mm lateral from bregma) using dental cement. A 3 mm craniotomy was drilled
916 (NSK UK Ltd.) in the centre of the headplate well and the dura removed. Virus was
917 then injected at a depth of 300 µm below the pia either as a single 750 nL injection at
918 200 nL/min or as ~5 injections of 150 nL virus at 50 nL/min spaced ~300 µm apart.
919 The pipette was left in the brain for 2 minutes after each injection. Following the final
920 retraction of the injection pipette, a two-tiered 4 mm/3 mm circle/circle chronic
921 window (UQG Optics cover-glass bonded with UV optical cement, NOR-61, Norland
922 Optical Adhesive) was press-fit into the craniotomy, sealed with cyanoacrylate
923 (Vetbond) and fixed in place with dental cement. After surgery, animals were
924 monitored and allowed to recover for at least 6 days during which they received water
925 and food *ad libitum*.
926
927 ## *Two-photon imaging*
928 For most experiments (**Figure 2** – **4**) two-photon imaging was performed using a
929 resonant scanning (30 Hz) microscope (Ultima II, Bruker Corporation) driven by
930 PrairieView and a Chameleon Ultra II laser (Coherent). For these experiments a
931 16x/0.8-NA water-immersion objective (Nikon) and an ETL (Optotune EL-10-30-TC,

Gardosoft driver) were used to collect 100 µm *z*-depth imaging volumes (4 planes, 33.3 µm spacing) with FOV sizes ranging from 600 x 600 µm to 850 x 850 µm (due to ETL magnification changes) at a constant image size of 512 x 512 pixels and plane-rate of ~7 Hz. The number of neurons recorded in each 2P psychometric curve experiment was 2809 ± 704, N = 11 sessions, 6 animals, 1 – 2 sessions each. For these experiments we used an orbital nosepiece that allows pitch, roll and yaw to vary in order to get the light-path through the objective orthogonal to the plane of the cranial window to ensure optimal imaging conditions (via the following procedure: https://github.com/llerussell/MONPangle). Due to inconsistencies in the window position across days and inaccuracies in the orbital nosepiece positioning and readout, pitch, roll and yaw can vary significantly across days making it difficult to image the same planes, and thus same neurons, in a volume. We therefore chose to reacquire new volumes each day and target new sets of neurons. For some experiments (some in **Figure 1,** some **Supplementary Figures**) two-photon imaging was performed using a resonant scanning (30 Hz) microscope (custom-build, Bruker Corporation) driven by PrairieView and a Chameleon Ultra II laser (Coherent). For these experiments a 25x/0.95-NA water-immersion objective (Leica) and an ETL (Optotune EL-10-30-TC, Optotune driver) were used to collect either 490 x 490 x 100 µm imaging volumes (4 planes with ~33 µm spacing) at an image size of 512 x 512 pixels and plane-rate of ~7 Hz or single planes of FOV size 490 x 490 µm and image size 512 x 512 pixels at 30 Hz. GCaMP6s was imaged at 920 nm and mCherry was imaged at 765 nm. For functional volumetric GCaMP6s imaging at 920nm the maximum power on sample was 50 mW at the shallowest plane (~130 µm below pia), increasing linearly to 80 mW at the deepest plane (~230 µm) to maintain image quality

across the volume (maximum 120 mins total duration per experiment). For single plane GCaMP6s imaging a maximum power of 50 mW was used for all depths. Power on sample for mCherry/mRuby2/mScarlett (conjugated to C1V1), imaged at 765 nm, was 50 – 100 mW (maximum 1 minute continuous duration).

Two-photon optogenetic stimulation

Two-photon photostimulation was carried out using a femto-second pulsed laser at 1030 nm (Satsuma, Amplitude Systèmes, 2 MHz rep-rate, 20 W). The single laser beam was split via a reflective spatial light modulator (SLM) (7.68 x 7.68 mm active area, 512 x 512 pixels, optimized for 1064nm, OverDrive Plus SLM, Meadowlark Optics/Boulder Nonlinear Systems) which was installed in-line of the photostimulation path. Phase masks used to generate beamlet patterns at the focal plane were calculated from photostimulation target *xy* co-ordinates centred on cell-bodies of interest via the weighted Gerchberg-Saxton (GS) algorithm. These targets were weighted according to their location relative to the centre of the SLM's addressable FOV to compensate for the decrease in diffraction efficiency when directing beamlets to peripheral positions. The transformation between SLM co-ordinates and imaging pixel co-ordinates was mapped by burning arbitrary spots in the FOV and calculating an affine transformation between SLM pixel targets and burn targets imaged in 2P. Calibration routines are available here https://github.com/llerussell/SLMTransformMaker3D. Spiral patterns were generated by moving all beamlets simultaneously with a pair of galvanometer mirrors. Each spiral consisted of 3 rotations (i.e. from centre to edge of spiral), 10 µm diameter, 20 ms duration. Powers were adjusted to maintain 6 mW per target neuron. Due to

constraints on the total power output of the photostimulation laser to stimulate 200 neurons we randomly divided all targeted neurons into 2 groups of 100 which were stimulated as an alternating pair such that each group of 100 neurons was stimulated 10 times at 20 Hz (stimulate first group with 20 ms spiral, 5 ms inter-spiral interval, stimulate second group with 20 ms spiral, 5 ms inter-spiral interval, then return to first group and repeat a further 9 times; each group therefore receives 10 x 20 ms spirals with an effective inter-spiral interval, for that group, of 25 ms; ~500 ms total stimulus duration). To stimulate 100 neurons or less all neurons were stimulated simultaneously 10 times at 40 Hz (10 x 20 ms spirals, 5 ms inter-spiral interval, ~250 ms stimulus duration). From our previous work we expect a single spiral to produce ~1 action potential (Packer et al., 2015), thus all 2P stimuli should drive ~10 action potentials. Spiral timing and positioning protocols were generated by a custom MATLAB software suite called Naparm and executed by the photostimulation modules of the microscope software (PrairieView, Bruker Corporation), the MeadowLark SLM software and our synchronisation software (PackIO; see below).

One-photon optogenetic stimulation

For Phase 1 of behavioural training an amber LED (590 nm peak wavelength, ThorLabs M590D2) was fixed to a manipulatable arm (DTI clamp, RS components) and press-fit onto the chronic window surface. For subsequent training phases an amber LED (595 nm peak wavelength, ThorLabs M595L3) was mounted in the lightpath above the two-photon microscope objective. LED powers ranged from 0.02 mW - 10 mW. Power and timing of LED photostimulation was controlled by custom MATLAB software through National Instruments data acquisition cards (NI USB-

6351, NI USB-6211). All LED stimuli consisted of 5 x 20 ms pulses at 25 Hz of varying powers ranging from 10 – 0.02 mW (measured with PM100A power meter/S130C photodiode sensor, ThorLabs). We did not record activity while animals were learning to detect 1P photostimulation. We thus roughly estimate the spatial extent, number and rate of spiking evoked by pulses of this power, duration and frequency from the literature as follows. Blue light emitted from an LED shows a ~1 mm HWHM spatial spread in brain tissue and ChR2-based pyramidal neuron activation shows relatively little variation with depth (Huber et al., 2008), though this could be due to activation of superficial processes. C-fos expression also shows that recruitment by optic fibre stimulation of ChR2-expressing pyramidal neurons is limited to a ~0.7 mm^3 volume around the fibre tip (Gradinaru, Mogri, Thompson, Henderson, & Deisseroth, 2009). Blue and orange light photoactivation of PV interneurons with moderate powers (10 mW) show similar lateral recruitment (~0.5 – 0.8 mm HWHM) (Li et al., 2019), as does orange light Arch-mediate inhibition of pyramidal neurons (~0.5 – 0.8 mm HWHM) (Babl, Rummell, & Sigurdsson, 2019). However, high power photobleaching indicates that orange light excitation has a two-fold larger lateral extent than blue light and extends across all cortical layers to form a roughly symmetric excitation volume (Li et al., 2019), as would be expected from the reduced scattering of red-shifted light in cortical tissue (Helmchen & Denk, 2005). Given the fact that we are using an LED instead of a fibre and we are using orange light, we suggest that our excitation volume has a roughly two-fold larger lateral extent than that measured for a blue LED (Huber et al., 2008) and for orange light from a fibre/laser (Babl et al., 2019; Li et al., 2019) and is roughly symmetric. With respect to stimulus frequency, C1V1 should be able to faithfully follow a 25 Hz pulse train with ~90% fidelity though our stimulus pulses

are longer than those used with C1V1 (Yizhar et al., 2011) or ChR2 (Huber et al., 2008). With respect to pulse duration, for ChR2 a doubling in pulse duration roughly corresponds to a doubling in spike probability (Histed & Maunsell, 2014) suggesting that our ten-fold increase in pulse duration relative to the literature (Yizhar et al., 2011) might correspond to a ten-fold increase in the number of evoked spikes. Though this might be an overestimate for C1V1 as it has slower channel kinetics than ChR2 (Yizhar et al., 2011), it is certainly likely we are driving multiple spikes with each pulse with an upper limit defined by C1V1's inability to reliably follow trains of > 50 Hz. These data, in combination with our observation that < 25% of neurons express C1V1 (**Figure 1 – figure supplement 1f**) and ~50% of C1V1-expressing neurons are photoactivatable, suggest that our initial training powers (10 mW) will drive > 5 spikes at between 25 – 50 Hz in ~10% of pyramidal neurons in a 1.5 – 2 mm^3 volume. The lowest powers at the behavioural threshold will likely recruit progressively fewer neurons over smaller cortical volumes with longer spike latencies (Huber et al., 2008) and these neurons will fire fewer spikes per pulse with lower reliability (Histed & Maunsell, 2014).

Synchronisation

For synchronisation of imaging frames, photostimulation spirals, sensory stimulation epochs and behavioural trial data during experiments, analogue triggers and waveforms were recorded with National Instruments DAQ cards controlled by PackIO (Watson, Yuste, & Packer, 2016). Behavioural trial timing and licking response contingency analyses were done online by an Arduino Mega microcontroller board controlled by PyBehaviour (Russell et al., 2019).

Online design and execution of photostimulation protocols

To quickly design photostimulation ensembles online during experiments we used our custom controls software, Naparm (https://github.com/llerussell/Naparm). Briefly, this software allows users to import images of neuronal populations of interest (i.e. C1V1 expression images) and semi-automatically detect the *xyz* centroids of neuron bodies. These potential target neurons can then be divided into different stimulation groups and the photostimulation protocol defined. Once defined this software saves out all files necessary to synchronise the microscope, lasers, SLM and master clock software PackIO. For behavioural training experiments we combined this with our custom two-photon behavioural training software (TPBS) which allows users to import a directory of phasemasks and associate one or more phasemasks with trial types that can be read into and executed by our behavioural control software PyBehaviour (see below).

Behavioural training

Training began ~21 days post virus injection/window installation. Animals were water-restricted to 85 – 90% of their pre-training weight throughout the training period and their weight was monitored daily. The majority of each animal's daily water was consumed during behavioural training, topped-up if necessary with additional water post-training. Early one-photon training sessions took place in closed, soundproofed and unlit behavioural training boxes. During training, animals were head-fixed via their headplates and housed in Perspex tubes. A metal reward delivery spout, connected to an electronic lickometer circuit, was positioned within easy reach of the mouse's

tongue to record licks and deliver sugar-water rewards (5 µL, 10% sucrose v/v). Mice were adapted to this procedure over a day or two during which rewards were randomly delivered manually. All subsequent behavioural training was controlled by Arduino-based behavioural control software written in Python (PyBehaviour: https://github.com/llerussell/PyBehaviour). This software acted as the master clock, dictating the sequence and timing of behavioural trials, and recording and scoring licking behaviour. The power and temporal characteristics of LED photostimulation patterns were controlled by custom-written MATLAB software and National Instruments hardware (see "One-photon optogenetic stimulation" section). PyBehaviour controlled the output of this LED software via TTL pulses. At the beginning of each behavioural session the LED, mounted on a manipulatable arm (DTI clamp, RS components), was press-fit onto the surface of the chronic window. This region was then sealed with tape to minimize direct visual stimulation. Throughout all training sessions and phases the basic task structure was the same. Trials were triggered after animals withheld licking for 7 ± 3 s. Each trial consisted of a response window during which licking behaviour was scored, followed by a 5 s post-stimulus period. The response window lasted for 2 s in training Phase 1 and 1 s in training Phase 2 and 3, though for all behavioural analyses a 1 s window beginning at 0.15 s (0.15 – 1.15 s) post stimulus was used to exclude unrealistically quick reaction times and ensure consistency of metrics across analyses of different phases (this window is indicated on all behavioural plots). Sessions consisted of 100 – 400 trials of two types: go trials and catch trials. During go trials, some form of optogenetic stimulus was delivered, and animals were required to lick. Licking was scored as a hit and was rewarded by delivery of a sugar water reward (see above), non-licking

was scored as a miss and was unpunished. During catch trials, no stimulus was delivered and animals were required not to lick. Licking was scored as a false alarm, and was unpunished, non-licking was scored as a correct reject and was unrewarded. Neither of these two trial types were cued. Stimulus trials therefore test an animal's detection rate on particular stimuli and catch trials assess an animal's chance response rate. All trial types were pseudorandomly interleaved with a 3 trial upper limit on consecutive trials of the same type. During the first few sessions go trials were auto-rewarded 500 ms following the stimulus to encourage learning. As soon as animals reliably began licking in anticipation of the auto-reward it was turned off for all subsequent sessions. Animals were initially trained on Phase 1 to detect 10 mW. Performance was then manually assessed and the LED power was dropped by half multiple times within a session until animals could detect very low powers (0.05 mW). This usually took ~4 days. Animals then underwent psychometric curve sessions to assess their performance on the lowest LED powers (0.1, 0.08, 0.06, 0.04, 0.02 mW). Animals that could detect 0.1 mW stimuli with d-prime > 1 were eligible for transitioning to subsequent experiments (all animals tested achieved this criterion), however only the subset with clearest expression were transitioned to two-photon training phases to allow for sufficient training time on the all-optical system. These animals were transitioned to Phases 2 and 3 where two-photon stimuli were introduced. For these phases animals were head-fixed under the microscope on a cylindrical treadmill and the LED was directed through the light-path. Note we expect that animals could be trained from the offset under the microscope using the LED through the objective, however we opted against this due to time constraints on the all-optical system. An objective well/baffle was used to minimize light leakage. A

white-noise mask was played continuously throughout all subsequent training sessions to mask any auditory cues emitted by the galvos during photostimulation. Following selection of neurons for 2P photostimulation (see below) animals began training on Phase 2 where 1P and 2P optogenetic stimuli were interleaved along with catch trials in equal proportions. Initial 2P stimuli targeted 200 neurons and animals required ~1 day to reach good performance. Once animals could reliably detect 200 neurons, we interleaved 2P stimuli targeting a 100 neuron subset of the original 200 neurons. Once animals could reliably detect 100 neurons they were transitioned to Phase 3 where only 2P optogenetic stimulation of those 100 neurons and catch trials were delivered in equal proportions. On some trials in a few Phase 2/3 sessions auto-rewards were delivered at 0.5, 1 or 1.5 s following the stimulus (depending on the animal's reaction time) to encourage transition to detecting two-photon stimuli. These trials were scored as hits if the animal licked before the auto-reward, otherwise they were conservatively scored as misses (even if they fell within the response window) (see *Behavioural data analysis* below). For the above phases we used the same FOV across days but didn't specifically target the same neurons (see *Two-photon imaging* section of *Methods*) except in the "Same" condition in **Figure 2 – figure supplement 3**. Following this, a subset of animals were transitioned to 2P psychometric curve sessions with the remaining subset used for other experiments not reported here (for details, see: https://discovery.ucl.ac.uk/id/eprint/10095170/). During two-photon psychometric curve sessions trial types were pseudorandomly interleaved to ensure an even distribution of trial types across the behavioural session. Trial-type ratios were as follows: 15% catch trials, 15% "easy" 200 neuron trials, 70% trials stimulating smaller numbers of neurons (~12% each for 100, 75, 50, 25, 10 and 5

neurons). All psychometric curve sessions had an initial 10-trial buffer of “easy” 200 neurons trials to allow the animal to warm up. Multiple FOVs were tested in each animal, but each FOV was tested only once. Mice were eligible for transfer from Phase 2 to Phase 3 and from Phase 3 to psychometric curve sessions once they’d achieved d-prime > 1 for detecting 2P stimulation of 200 neurons. This did not always happen immediately due to time constraints on the all-optical system. If transition was delayed, in the intervening days mice were often trained to keep them familiar with the task. Summary of *N* for behavioural training: 26 mice did 1P training and of these 18 mice did both high power and low power 1P psychometric curves and 4 did just high power and 4 did just low power (these tended to be the best mice which were rushed to later training phases) (**Figure 1**, **Figure 1 – figure supplement 3**). From these, 12 mice with the clearest expression were transitioned onto 2P training (**Figure 1**, **Figure 1 – figure supplement 5**) and from these 6 mice were used for 2P psychometric curve sessions (**Figures 2 – 4**) and the remaining 6 were used for other 2P experiments not reported here.

Image processing

All offline data were analysed using custom software and toolboxes in MATLAB. PackIO data acquired during experiments was used to synchronise imaging frames with photostimulation and behavioural epochs. To avoid the large imaging artefacts caused by 2P photostimulation, we implemented a “stimulus artefact exclusion epoch” whereby all imaging frames acquired during photostimulation periods were excluded from all processing, analyses and plotting. These is described below. Since it takes 500 ms to stimulate 200 neurons (see *Two-photon optogenetic stimulation*

above), and since we wanted the amount of data (volume scans) excluded to be the same for all trial types (to facilitate comparison across them), we need to exclude at least 0 - 500 ms of imaging data post-stimulus onset on all trials (even though stimuli targeting <= 100 neurons take ~250 ms). In reality we actually have to remove a slightly longer period of imaging data (5 full volumes peri-stimulus; ~750 ms of imaging data) due to how the photostimulus epoch overlaps with the time-course of imaging volume acquisition. We explain this below. Photostimulation onsets are not synchronized to begin at the same time as the first plane in a given volume (i.e. they can occur during any plane within a volume). Our volumes contain 4 planes and if any plane is corrupted by the stimulus artifact then that volume must be discarded. Our imaging frame (plane) rate is 26.8 Hz and, since we acquire 4 planes per volume, our volume rate is 6.69 Hz. This means it takes ~150 ms to acquire the four planes constituting one volume. Thus, given that the maximum photostimulus duration is ~500 ms (see above), even if photostimulus onset was synchronized to always begin with the first plane of a given volume, the minimum number of volumes we could discard post-stimulus onset would be 4 (4 * 150 ms = 600 ms). However, since the photostimulus onset can occur at any time during a volume acquisition, it can also occur during the last plane of the volume immediately preceding the stimulus onset. This volume must also be discarded in addition to the subsequent 4 volumes which must also be discarded as they will also all contain at least one frame corrupted by the photostimulus. This means that on some trials 5 volumes will need to be discarded (~750 ms of imaging data). Since we always want to discard the same amount of imaging data across trials to facilitate comparison between them, we have to exclude this maximum value of 5 volumes (~750 ms) peri-stimulus on all trials.

Since most reaction times occur before this (560 ± 150 ms reaction time for 2P 200 neuron trials) we are unable to analyse data in the absence of response-related activity (licking, whisking, facial movements). Thus we had to control for this in our subsequent analyses (see *Neurometric curve analyses* section of *Methods*). Imaging time-series were registered and segmented into ROIs using the Python version of Suite2P (Pachitariu et al., 2016). ROIs were manually curated. Neuropil subtracted neuron traces were calculated as:

$$F_{cell} = F_{ROI} - c * F_{neuropil}$$

where the neuropil subtraction coefficient c was estimated separately for each ROI by robust regression between F_{ROI} and $F_{neuropil}$ (T.-W. Chen et al., 2013). For this estimation process, all photostimulation epochs were excluded and F_{ROI} and $F_{neuropil}$ were downsampled by a factor of 10 (to 0.7 Hz). Coefficients were post-hoc bounded between 0.5 and 1 and any coefficients that could not be reliably estimated were set to the median of all reliably estimated coefficients in that dataset (usually ~0.7). Once $F_{neuropil}$ had been subtracted from F_{ROI}, F_{cell} was then re-baselined to the 33rd percentile of F_{ROI} values to ensure accuracy in subsequent $\Delta F/F_0$ and $\Delta F/\sigma F$ calculations. Neuropil subtraction had a small but significant effect on trial-wise response amplitude, however we observed a large fraction of negative responses even without neuropil subtraction and response classifications (activated, suppressed and no response) were largely consistent between raw and neuropil subtracted trial-wise responses (**Figure 3 – figure supplement 2**). To detect neurons expressing C1V1 offline post experiment (**Figure 1 – figure supplement 1**) we used the Cellpose algorithm followed by manual curation (Stringer, Wang, Michaelos, & Pachitariu, 2020) using the *cytoplasm* model with diameter = 15 μm.

Neuronal response analysis

For trial-wise analyses, F_{cell} traces were divided into epochs triggered on photostimulation onset for stimulus trials and catch trial response window onset for catch trials (stimulus-triggered averages – STAs). All STAs had a 1 s baseline period. STAs for each trial were converted to $\Delta F/\sigma F$ by subtracting the baseline mean from the STA trace (ΔF) and dividing the result by the baseline standard deviation (σF). Responses on both photostimulation and catch trials were quantified as the average response ~0.7 – 1 s following photostimulus onset so as to avoid photostimulation artefacts (see *Image processing* section above for full description of the stimulus artefact exclusion epoch). For our analyses we wanted to be able to assess neural responses on single trials as our procedure for mitigating potential lick response artefacts requires trial-wise analyses. Moreover we cannot assume that the same background neurons will be recruited by photostimulation of a given set of target neurons on each trial due to the high variability and low probability of synaptic transmission (London et al., 2010). This therefore precludes the use of a standard statistical test comparing a neuron's stimulus trial response distribution to its catch trial response distribution to assess responsivity. We therefore sought to define activation and suppression thresholds that we could apply to responses on individual trials. We used each neuron's response distribution on correct reject (CR) catch trials, when no stimulus or lick response occurred, to define its separate activation and suppression thresholds (since activation and suppression are readout differently by calcium indicators; see Otis et al., 2017) (**Figure 2 – figure supplement 1b – g**). For each neuron we defined its activation threshold as:

$$Threshold_{activation} = Mean_{catch} + S.D._{catch} * Scaling\ Factor_{activation}$$

and its suppression threshold as:

$$Threshold_{suppression} = Mean_{catch} - S.D._{catch} * Scaling\ Factor_{suppression}$$

Where $Mean_{catch}$ and $S.D._{catch}$ are the mean and standard deviation of the distribution of CR catch trial responses respectively (**Figure 2 – figure supplement 1b**). We estimated the activation and suppression scaling factors separately for each session (i.e. each session will have a single activation and single suppression scaling factor applied to all neurons), and separately from each other, using a cross-validated empirical procedure where the objective was to ensure that only 5% of neurons were activated and 5% of neurons were suppressed on catch trials, effectively fixing the false positive (FP) rate at 5% for each session (**Figure 2 – figure supplement 1c**). For data recorded on each session and for each response type (i.e. in the following example for activation) we swept through a series of potential activation scaling factors (range 1 – 3 in increments of 0.1). At each scaling factor we ran 10,000 permutations of an 80:20 train:test split over CR catch trials. On each permutation we used the training CR catch trials to define each neuron's activation threshold (using the equation above with the current scaling factor) and then used these thresholds to calculate the proportion of neurons crossing their threshold on each testing CR catch trial and averaged this across trials to get the average proportion activated across testing trials. We then took the median proportion activated across all permuted train:test splits. We plot this median value for all scaling factors resulting in a distinct curve for each session (**Figure 2 – figure supplement 1d,e** data points). We fit each session's curve with cubic interpolation (**Figure 2 – figure supplement 1d,e** curves) and then used this to infer the s.d. scaling factor that yields a 5% FP

rate across permuted train:test splits for each session and for each response type (**Figure 2 – figure supplement 1d,e** box plots). These scaling factors can then be used in conjunction with the full CR catch response distribution as described in the equations above to define an activation threshold and a suppression threshold for each neuron in each session. For subsequent analyses, to quantify the number of target neurons activated we simply calculated the number of neurons in target zones that passed their activation threshold on each trial and averaged this across trials of the same type. To quantify P(Activated) and P(Suppressed) in all neurons (**Figure 3a – c**) we divided the total number of activated or suppressed neurons on each trial, irrespective of whether they were in target zones or not, by the total number of neurons (targets and background) and averaged across trials of the same type. To quantify P(Activated) and P(Suppressed) in background neurons (**Figure 3d,e**) we divided the total number of activated or suppressed background neurons on each trial (target neurons excluded) and divided this by the total number of neurons (targets and background) and averaged across trials of the same type. Note that all neural response data in **Figure 3** and **4** are hit:miss matched (see below *Hit:miss matching*).

Classifying target neurons and background neurons

Since the spatial resolution of two-photon photostimulation is not perfect we defined conservative 3D photostimulation zones around each targeted location in the imaging volume. These had a 10 μm radius, which is ~2x the lateral resolution of our 2P photostimulation (**Figure 1 – figure supplement 2a**), and extended through the entire axial range of the volume (99 μm) (**Figure 2 – figure supplement 1a**). Thus we considered all ROIs with a centroid <= 10 μm from any target location to be a potential

target neuron, irrespective of their axial displacement relative to the target location. ROIs outside these zones were considered background neurons. Since these exclusion zones are conservative, many neurons fell within them (**Figure 2 – figure supplement 1h**), though only a fraction were responsive to photostimulation (**Figure 2 – figure supplement 1i,j** 0.18 ± 0.1 fraction activated across neurons within target zones, averaged across trial types, N = 11 sessions, 6 animals, 1 – 2 sessions each). Nevertheless this fraction was far above the fraction responsive on catch trials (**Figure 2 – figure supplement 1j** stimulus trials: 0.18 ± 0.1 vs catch trials: 0.04 ± 0.01 averaged across trial types $P = 4.92 \times 10^{-4}$ paired t-test, N = 11 sessions, 6 animals, 1 – 2 sessions each) and resulted in numbers of activated target neurons that were 0.46 ± 0.20 times that of the number of target zones (averaged across trial types) and decreased with decreasing number of zones as intended (**Figure 2 – figure supplement 1i**).

Hit:miss matching

We are interested in how the number of target neurons activated modulates responses in the local network (**Figure 3,4**). The number of target neurons activated increases the probability of licking (**Figure 2**). Licking itself activates responses in the background network irrespective of photostimulation of target neurons (**Figure 3 – figure supplement 1**). Therefore we sought to minimize the chance that any modulation in background network activity we observe in response to changes in the number of target neurons activated is due to increased recruitment of licking, and lick-evoked activity, by more salient target stimulation. Thus we fixed a 50:50 ratio of hits:misses (lick:no-lick trials) on every trial type (number of target zones) (**Figure 3 –**

figure supplement 1o). In this way we can ensure that the proportion of lick corruption in hard to detect stimuli (5 target zones), which mainly result in misses, is similar to that of easy to detect stimuli (100 target zones), which mainly result in hits. Below we describe this procedure using the least salient trial type (5 target zones) of a given session as an example. First we count the number of trials of this trial type with the minority response type which, in this case, is hits since animals rarely detect 5 target site stimulations. This number is logged as the number of trials to match with trials of the majority response type which, in this case, is misses (as animals often fail to detect 5 target zone stimulations). We then run 100 permutations where we take all the minority trial type trials (hits) and random resamples, of the same number, of the majority trial type trials (misses). For example if there were 2 hits and 15 misses overall then on each permutation we would take the same 2 hit trials and a random 2 miss trials (sampled from the 15 misses) and calculate our neural metrics across these trials. We then average these metrics across all permutations to get final values reported in network response figures. This procedure was not possible for a few trial types as there only trials of one response type (i.e. only hits or only misses). These trial types were excluded from analysis.

Lick analysis

To correlate spontaneous fluorescence traces with spontaneous licking we removed photostimulation trial periods (0 – 4 s post-stimulus) and analysed the remaining periods of lick and fluorescence traces spanning the time between the first and last spontaneous lick. We smoothed the lick trace with a Gaussian kernel (sigma = 0.5 s) and calculated the Pearson's correlation coefficient (and associated *P*-values)

between the smoothed lick trace and each neuron's fluorescence trace. To detect spontaneous lick bouts for lick-triggered STA analysis we excluded all trial periods (0 – 4 s post-stimulus onset) and found groups of at least 3 licks that occurred with a minimum 0.5 s inter-lick interval that were separated from preceding licks by at least 1 s.

Behavioural data analysis

All behavioural trials with <= 0.15 s reaction time were excluded from analysis. Response windows for analysis of all Phases extended for 1 s following this period (i.e. 0.15 – 1.15 s post-stimulus onset). This is indicated in all raster plots. Behavioural trials during imaging experiments which occurred so close to the beginning/end of imaging epochs that the requisite analysis windows were truncated were excluded from analyses. For calculations of reaction time standard deviation we only included sessions where all trial types had responses on > 2 trials. For sessions including some auto-rewarded trials (early Phase 1 and Phase 2 sessions) we scored auto-rewarded trials with response times < $t_{auto\text{-}reward}$ + 0.15 s as hits and scored auto-rewarded trials with responses after this time, or no response at all, as misses. The basic behavioural response metric used throughout was the P(response) for each stimulation type, calculated as:

$$P(response) = \frac{n_{lick}}{n_{lick} + n_{no\ lick}}$$

Where n_{lick} and $n_{no\ lick}$ are the number of trials where the animal licked and did not lick respectively. For some analyses the catch trial $P(response)$ was subtracted from the $P(response)$ on go trials. This has been indicated where relevant as "Catch subtracted". For some behavioural analyses where learning relative to our

behavioural response criterion (d-prime > 1) was relevant we report performance in terms of d-prime which we computed as:

$$norminv(Hit\ rate) - norminv(False\ alarm\ rate)$$

where $norminv$ is the inverse of the normal cumulative distribution function. We corrected for hit rates/false alarm rates of 1 and 0 according to Macmillan & Kaplan, 1985.

Psychometric curve fitting

We used the MATLAB *psignifit* toolbox (Schütt, Harmeling, Macke, & Wichmann, 2016; https://github.com/wichmann-lab/psignifit) to fit log-normal beta-binomial psychometric curves of the following form:

$$\psi(x) = \gamma + (1 + \lambda - \gamma)\, S(x; m, w)$$

$$S(x; m, w) = \Phi\left(C \frac{\log(x) - m}{w}\right)$$

$$C = \Phi^{-1}(0.95) - \Phi^{-1}(0.05)$$

For which we fit the width (w) and threshold (m) and fixed the lapse rate (λ) and guess rate (γ). $\log$ is the natural logarithm. The lapse rate was fixed as 1 – max(P(Lick)$_{Stim}$) where max(P(Lick)$_{Stim}$) is the maximum response rate to any stimulus trial-type (this was always either 100 or 200 target zone stimulation trials). The guess rate was fixed as the catch trial response rate. Only go (stimulus) trials were used for fitting. For fits aggregating across sessions (**Figure 2d** red curve) the lapse rate and guess rate were fixed using values averaged across sessions. For fits during train/test permutations (**Figure 4d**) the lapse rate and guess rate were fixed at values computed from all data and fixed for all permutations whereas the width and threshold were re-fit for every permutation.

Data exclusion criteria

Some animals were removed from analysis of reaction time variability in Phase 1 behavioural sessions as they did not show enough responses in catch trials to compute the standard deviation of reaction time (**Figure 1f**). Some trial types for some sessions were excluded from **Figure 3** and **4** as the procedure required to match the number of Hit and Miss trials resulted in no trials of certain trial types (see *Hit:miss matching* section of the *Methods*). Some trial types were removed from network response analyses (**Figure 3**) as the number of targets activated was < 1 on average. No data were removed as outliers. Trials with reaction times <= 0.15 s were excluded from all figures (except for illustrative purposes in **Figure 1 – figure supplement 3b**) and analyses as this was deemed to be too quick to be a reaction to the stimulus. In the latter half of each training session if the response probability on the easiest trial type for that session fell below 0.7 (calculated in a 10-trial sliding window centred on each trial) the animal was deemed to be "sated" and trials beyond this point were excluded from all analyses.

Statistical procedures

No statistical procedures were used to estimate sample sizes. Appropriate sample sizes were estimated from previous experiments of a similar type. Experiments were not randomized and experimenters were not blinded with respect to experiment and outcome. All data were analysed with custom routines and toolboxes in MATLAB except for psychometric curve fitting which was done using the MATLAB *psignifit* toolbox (see *Psychometric curve fitting* section of *Methods*). All error bars are given

as mean ± s.e.m and all values in the text are mean ± s.d. unless otherwise stated. All boxplots show median and 25th/75th percentile boxes with whiskers extending to the most extreme data not considered outliers (outliers defined as data $> q_3 + 1.5 \times (q_3 - q_1)$ or $< q_3 - 1.5 \times (q_3 - q_1)$ where q_3 and q_1 are the 75th and 25th percentiles respectively). Regression coefficients and psychometric function parameters are reported ± 95% confidence intervals where appropriate. Datasets of $N > 8$ were tested for normality with D'Agostino-Pearson's K2 test and analysed according to the result. Datasets of $N \leq 8$ were analysed as non-parametric. Multiple comparisons were corrected using the Bonferroni correction. All tests were two-tailed. Relevant tests are reported in the text.

Data availability

Import, processing, analysis and figure code is available on Github (Dalgleish, 2020; https://github.com/alloptical/Dalgleish-eLife-2020) for use with analysed data (https://doi.org/10.6084/m9.figshare.13135505) and/or unprocessed behavioural session data (https://doi.org/10.6084/m9.figshare.13128950). Raw calcium imaging movies are ~1TB in size and are thus available upon reasonable request.

Acknowledgements

We thank Mehmet Fişek for useful discussions about experiments and analysis; Selmaan Chettih and Christopher Harvey for developing and sharing the somatically-restricted C1V1 opsin; Soyon Chun and Agnieszka Jucht for mouse breeding; Sarolta Gabulya for behavioural training during pilot experiments; and Bruker Corporation for technical support. This work was supported by grants from the Wellcome Trust, Gatsby Charitable Foundation, ERC, MRC and the BBSRC.

Competing interests

The authors declare no competing interests.

References

Abraham, N. M., Spors, H., Carleton, A., Margrie, T. W., Kuner, T., & Schaefer, A. T. (2004). Maintaining Accuracy at the Expense of Speed: Stimulus Similarity Defines Odor Discrimination Time in Mice. *Neuron*, *44*, 865–876.

Alejandre-García, T., Kim, S., Pérez-Ortega, J., & Yuste, R. (2020). Intrinsic excitability mechanisms of neuronal ensemble formation. *BioRxiv*, 1–43.

Amit, D., Gutfreund, H., & Sompolinsky, H. (1985a). Spin-glass models of neural networks. *Physical Review A*, *32*(2), 1007–1018.

Amit, D., Gutfreund, H., & Sompolinsky, H. (1985b). Storing infinite numbers of patterns in a Spin-Glass model of neural networks. *Physical Review Letters*, *55*(14), 1530–1533.

Andermann, M. L., Kerlin, A. M., & Reid, R. C. (2010). Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. *Front Cellular Neuroscience*, *4*(March), 3. https://doi.org/10.3389/fncel.2010.00003

Attwell, D., & Laughlin, S. B. (2001). An Energy Budget for Signaling in the Grey Matter of the Brain. *Journal of Cerebral Blood Flow and Metabolism*, *21*, 1133–1145.

Babl, S. S., Rummell, B. P., & Sigurdsson, T. (2019). The Spatial Extent of Optogenetic Silencing in Transgenic Mice Expressing Channelrhodopsin in Inhibitory Interneurons. *Cell Reports*, *29*, 1381–1395. https://doi.org/10.1016/j.celrep.2019.09.049

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*, 371–394.

Barth, A. L., & Poulet, J. F. a. (2012). Experimental evidence for sparse firing in the neocortex. *Trends in Neurosciences*, *35*(6), 345–355. https://doi.org/10.1016/j.tins.2012.03.008

Baum, E. B., Moody, J., & Wilczek, F. (1988). Internal representations for associative memory. *Biological Cybernetics*, *228*, 217–228.

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bernander, O., Douglas, R., Martin, K., & Koch, C. (1991). Synaptic background activity influences spatiotemporal integration in single pyramidal cells. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(December), 11569–11573.

Bernardi, D., Doron, G., Brecht, M., & Lindner, B. (2020). A network model of the barrel cortex combined with a differentiator detector reproduces features of the behavioral response to single-neuron stimulation. *BioRxiv*, 1–35.

Bernardi, D., & Lindner, B. (2017). Optimal Detection of a Localized Perturbation in Random Networks of Integrate-and-Fire Neurons. *Physical Review Letters*, *118*(268301), 1–5. https://doi.org/10.1103/PhysRevLett.118.268301

Bernardi, D., & Lindner, B. (2019). Detecting single-cell stimulation in a large network of integrate-and-fire neurons. *Physical Review E*, *99*(032304), 1–18. https://doi.org/10.1103/PhysRevE.99.032304

Bialek, W. (2012). *Biophysics: Searching for Principles*. Princeton University Press. Retrieved

from https://books.google.co.uk/books?id=MLYJ5Rz3GLwC

Bock, D. D., Lee, W. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., … Reid, R. C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, *471*, 177–182. https://doi.org/10.1038/nature09802

Borg-Graham, L. J., Monier, C., & Frégnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, *7603*(1979), 619–623.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. a. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, *13*(1), 87–100. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8730992

Brunel, N. (2016). Is cortical connectivity optimized for storing information? *Nature Neuroscience*, *19*(5). https://doi.org/10.1038/nn.4286

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

Cai, B., Billeh, Y. N., Chettih, S. N., Harvey, C. D., Koch, C., Arkhipov, A., & Mihalas, S. (2020). Modeling robust and efficient coding in the mouse primary visual cortex using computational perturbations. *BioRxiv*, 1–37.

Carandini, M., & Churchland, A. K. (2013). Probing perceptual decisions in rodents. *Nature Neuroscience*, *16*(7), 824–831. https://doi.org/10.1038/nn.3410

Carbon, C.-C. (2014). Understanding human perception by human-made illusions. *Frontiers in Human Neuroscience*, *8*(July), 1–6. https://doi.org/10.3389/fnhum.2014.00566

Cardin, J. A., Kumbhani, R. D., Contreras, D., & Palmer, L. A. (2010). Cellular Mechanisms of Temporal Sensitivity in Visual Cortex Neurons. *The Journal of Neuroscience*, *30*(10), 3652–3662. https://doi.org/10.1523/JNEUROSCI.5279-09.2010

Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A., Carrillo-reid, L., Han, S., … Yuste, R. (2019). Controlling Visually Guided Behavior by Holographic Recalling of Cortical Ensembles. *Cell*, *178*, 1–11. https://doi.org/10.1016/j.cell.2019.05.045

Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D. S., & Yuste, R. (2016). Imprinting and recalling cortical ensembles. *Science*, *353*, 691–694. https://doi.org/10.1126/science.aaf7560

Cassidy, C. M., Balsam, P. D., Rosengard, R. J., Slifstein, M., Daw, N. D., Abi-dargham, A., & Horga, G. (2018). A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. *Current Biology*, *28*(4), 503–514. https://doi.org/10.1016/j.cub.2017.12.059

Ceballo, S., Bourg, J., Kempf, A., Piwkowska, Z., Daret, A., Pinson, P., … Bathellier, B. (2019). Cortical recruitment determines learning dynamics and strategy. *Nature Communications*, *10*, 1–12. https://doi.org/10.1038/s41467-019-09450-0

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

Chaudhury, S. (2010). Hallucinations: Clinical aspects and management. *Industrial Psychiatry Journal*, *19*(1), 5–12. https://doi.org/10.4103/0972-6748.77625

Chen, J. L., Carta, S., Soldado-Magraner, J., Schneider, B. L., & Helmchen, F. (2013).

Behaviour-dependent recruitment of long-range projection neurons in somatosensory cortex. *Nature*, *499*(7458), 336–340. https://doi.org/10.1038/nature12236

Chen, J. L., Margolis, D. J., Stankov, A., Sumanovski, L. T., Schneider, B. L., & Helmchen, F. (2015). Pathway-specific reorganization of projection neurons in somatosensory cortex during learning. *Nature Neuroscience*, *18*(8), 1101–1108. https://doi.org/10.1038/nn.4046

Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., … Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, *499*(7458), 295–300. https://doi.org/10.1038/nature12354

Chettih, S. N., & Harvey, C. D. (2019). Single-neuron perturbations reveal feature-specific competition in V1. *Nature*, *567*, 334–340. https://doi.org/10.1038/s41586-019-0997-6

Cohen, M. R., & Newsome, W. T. (2004). What electrical microstimulation has revealed about the neural basis of cognition. *Current Opinion in Neurobiology*, *14*, 169–177. https://doi.org/10.1016/j.conb.2004.03.016

Constantinople, C. M., & Bruno, R. M. (2013). Deep Cortical Layers Are Activated Directly by Thalamus. *Science*, *1591*(June), 1591–1594. https://doi.org/10.1126/science.1236425

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-day, B., Schmack, K., & Powers III, A. R. (2019). Hallucinations and Strong Priors. *Trends in Cognitive Sciences*, *23*(2), 114–127. https://doi.org/10.1016/j.tics.2018.12.001

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., … Mrsic-flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, *518*(00), 399–403. https://doi.org/10.1038/nature14182

Daie, K., Svoboda, K., & Druckmann, S. (2019). Targeted photostimulation uncovers circuit motifs supporting short-term memory. *BioRxiv*.

Dalgleish, H. W. P. (2020). Dalgleish-eLife-2020. *Github*. https://github.com/alloptical/Dalgleish-eLife-2020. 7e47864.

Denève, S., & Machens, C. K. (2016). Efficient codes and balanced networks. *Nature Neuroscience*, *19*(3), 375–382. https://doi.org/10.1038/nn.4243

Destexhe, A., & Pare, D. (1999). Impact of Network Activity on the Integrative Properties of Neocortical Pyramidal Neurons In Vivo. *Journal of Neurophysiology*, *81*(4), 1531–1547.

Destexhe, A., Rudolph, M., & Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nature Reviews Neuroscience*, *4*(September), 739–751. https://doi.org/10.1038/nrn1198

Doron, G., von Heimendahl, M., Schlattmann, P., Houweling, A. R., & Brecht, M. (2014). Spiking Irregularity and Frequency Modulate the Behavioral Report of Single-Neuron Stimulation. *Neuron*, *81*(3), 653–663. https://doi.org/10.1016/j.neuron.2013.11.032

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., & Suarez, H. H. (1995). Recurrent Excitation in Neocortical Circuits. *Science*, *269*(August), 981–986.

Eichenbaum, H., & Cohen, N. J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, *83*(4), 764–770. https://doi.org/10.1016/j.neuron.2014.07.032

Emiliani, V., Cohen, A. E., Deisseroth, K., & Hausser, M. (2015). All-Optical Interrogation of Neural Circuits. *The Journal of Neuroscience*, *35*(41), 13917–13926. https://doi.org/10.1523/JNEUROSCI.2916-15.2015

Fahle, M. (2002). Perceptual learning: specificity versus generalization. *Current Opinion in*

Neurobiology, *15*, 154–160. https://doi.org/10.1016/j.conb.2005.03.010

Feldmeyer, D., Brecht, M., Helmchen, F., Petersen, C. C. H., Poulet, J. F. A., Staiger, J. F., … Schwarz, C. (2013). Barrel cortex function. *Progress in Neurobiology*, *103*, 3–27. https://doi.org/10.1016/j.pneurobio.2012.11.002

Fino, E., & Yuste, R. (2011). Dense Inhibitory Connectivity in Neocortex. *Neuron*, *69*(6), 1188–1203. https://doi.org/10.1016/j.neuron.2011.02.025

Folli, V., Leonetti, M., & Ruocco, G. (2017). On the Maximum Storage Capacity of the Hopfield Model. *Frontiers in Computational Neuroscience*, *10*(January), 1–9. https://doi.org/10.3389/fncom.2016.00144

Friston, K. (2005). Hallucinations and perceptual inference. *Behavioral and Brain Sciences*, *28*(May 2020), 764–766.

Gerdjikov, T. V, Bergner, C. G., Stüttgen, M. C., Waiblinger, C., & Schwarz, C. (2010). Discrimination of Vibrotactile Stimuli in the Rat Whisker System: Behavior and Neurometrics. *Neuron*, *65*, 530–540. https://doi.org/10.1016/j.neuron.2010.02.007

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The Neural Basis of Perceptual Learning. *Neuron*, *31*, 681–697.

Gill, J. V, Lerman, G. M., Zhao, H., Stetler, B. J., Rinberg, D., & Shoham, S. (2020). Precise Holographic Manipulation of Olfactory Circuits Reveals Coding Features Determining Perceptual Detection. *Neuron*, *108*, 1–12. https://doi.org/10.1016/j.neuron.2020.07.034

Gradinaru, V., Mogri, M., Thompson, K. R., Henderson, J. M., & Deisseroth, K. (2009). Optical Deconstruction of Parkinsonian Neural Circuitry. *Science*, *324*(April), 354–360.

Gustafson, N. J., & Daw, N. D. (2011). Grid Cells, Place Cells, and Geodesic Generalization for Spatial Reinforcement Learning. *PLoS Computational Biology*, *7*(10), e1002235. https://doi.org/10.1371/journal.pcbi.1002235

Haider, B., Häusser, M., & Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature*, *493*(7430), 97–100. https://doi.org/10.1038/nature11665

Haider, B., & McCormick, D. A. (2009). Rapid Neocortical Dynamics: Cellular and Network Mechanisms. *Neuron*, *62*(2), 171–189. https://doi.org/10.1016/j.neuron.2009.04.008

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, *56*(1), 51–65.

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, *503*(7474), 51–58. https://doi.org/10.1038/nature12654

Helmchen, F., & Denk, W. (2005). Deep tissue two-photon microscopy. *Nature Methods*, *2*(12), 932–940. https://doi.org/10.1038/NMETH818

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frotiers in Human Neuroscience*, *3*(November), 1–11. https://doi.org/10.3389/neuro.09.031.2009

Herculano-Houzel, S., Collins, C. E., Wong, P., & Kaas, J. H. (2007). Cellular scaling rules for primate brains. *Proceedings of the National Academy of Sciences*, *104*(9), 3562–3567.

Herculano-Houzel, S., Mota, B., & Lent, R. (2006). Cellular scaling rules for rodent brains. *Proceedings of the National Academy of Sciences*, *103*(32), 12138–12143.

Hires, S. A., Gutnisky, D. A., Yu, J., O'Connor, D. H., & Svoboda, K. (2015). Low-noise encoding of active touch by layer 4 in the somatosensory cortex. *ELife*, *4*, e06619. https://doi.org/10.7554/eLife.06619

Histed, M. H., Carvalho, L. A., & Maunsell, J. H. R. (2012). Psychophysical measurement of

contrast sensitivity in the behaving mouse. *Journal of Neurophysiology*, *107*, 758–765. https://doi.org/10.1152/jn.00609.2011

Histed, M. H., & Maunsell, J. H. R. (2014). Cortical neural populations can guide behavior by integrating inputs linearly, independent of synchrony. *Proceedings of the National Academy of Sciences*, *111*(1), E178-87. https://doi.org/10.1073/pnas.1318750111

Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., … Mrsic-Flogel, T. D. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, *14*(8), 1045–1052. https://doi.org/10.1038/nn.2876

Holmgren, C., Harkany, T., Svennenfors, B., & Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *Journal of Physiology*, *551*(1), 139–153. https://doi.org/10.1113/jphysiol.2003.044784

Hong, Y. K., Lacefield, C. O., Rodgers, C. C., & 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., Hires, S. A., Zhang, Y., Huber, D., Petreanu, L., & Shepherd, G. M. G. (2011). Laminar Analysis of Excitatory Local Circuits in Vibrissal Motor and Sensory Cortical Areas. *PLoS Biology*, *9*(1), e10000572. https://doi.org/10.1371/journal.pbio.1000572

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

Houweling, A. R., & Brecht, M. (2008). Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, *451*(7174), 1–8. https://doi.org/10.1038/nature0

Huber, D., Gutnisky, D. A., Peron, S., O'Connor, D. H., Wiegert, J. S., Tian, L., … Svoboda, K. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature*, *484*(7395), 473–478. https://doi.org/10.1038/nature11039

Huber, D., Petreanu, L., Ghitani, N., Ranade, S., Hromádka, T., Mainen, Z., & Svoboda, K. (2008). Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. *Nature*, *451*(7174), 61–64. https://doi.org/10.1038/nature06445

Isaacson, J. S., & Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron*, *72*(2), 231–243. https://doi.org/10.1016/j.neuron.2011.09.027

Jazayeri, M., & Afraz, A. (2017). Navigating the Neural Space in Search of the Neural Code. *Neuron*, *93*(5), 1003–1014. https://doi.org/10.1016/j.neuron.2017.02.019

Jennings, J. H., Kim, C. K., Marshel, J. H., Raffiee, M., Ye, L., Quirin, S., … Deisseroth, K. (2019). Interacting neural ensembles in orbitofrontal cortex for social and feeding behaviour. *Nature*, *565*, 645–649. https://doi.org/10.1038/s41586-018-0866-8

Jouhanneau, J.-S., Kremkow, J., & Poulet, J. F. A. (2018). Single synaptic inputs drive high-precision action potentials in parvalbumin expressing GABA-ergic cortical neurons in vivo. *Nature Communications*, *9*(1), 1540. https://doi.org/10.1038/s41467-018-03995-2

Kanerva, P. (1993). Sparse Distributed Memory and Related Models. In M. H. Hassoun (Ed.), *Associative Neural Memories: Theory and Implementation* (pp. 50–76). Oxford University Press Inc.

Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A. L., Dhawale, A. K., … Ölveczky, B. P. (2015). Motor Cortex Is Required for Learning but Not for Executing a Motor Skill. *Neuron*, 800–812. https://doi.org/10.1016/j.neuron.2015.03.024

Kerlin, A. M., Andermann, M. L., Berezovskii, V. K., & Reid, R. C. (2010). Broadly Tuned

Response Properties of Diverse Inhibitory Neuron Subtypes in Mouse Visual Cortex. *Neuron*, *67*(5), 858–871. https://doi.org/10.1016/j.neuron.2010.08.002

Khan, A. G., Poort, J., Chadwick, A., Blot, A., Sahani, M., Mrsic-flogel, T. D., & Hofer, S. B. (2018). Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nature Neuroscience*, *21*(June). https://doi.org/10.1038/s41593-018-0143-z

Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S. B., & Mrsic-Flogel, T. D. (2013). The emergence of functional microcircuits in visual cortex. *Nature*, *496*(7443), 96–100. https://doi.org/10.1038/nature12015

Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. a, Sjöström, P. J., & Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, *473*(7345), 87–91. https://doi.org/10.1038/nature09880

Kumar, S., Soren, S., & Chaudhury, S. (2009). Hallucinations: Etiology and clinical implications. *Industrial Psychiatry Journal*, *18*(2), 119–126. https://doi.org/10.4103/0972-6748.62273

Kwan, A. C., & Dan, Y. (2012). Dissection of cortical microcircuits by single-neuron stimulation in vivo. *Current Biology*, *22*(16), 1459–1467. https://doi.org/10.1016/j.cub.2012.06.007

Kwon, S. E., Yang, H., Minamisawa, G., & O'Connor, D. H. (2016). Sensory and decision-related activity propagate in a cortical feedback loop during touch perception. *Nature Neuroscience*, *19*(9), 1243–1252. https://doi.org/10.1038/nn.4356

Law, C.-T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience*, *11*(4), 505–513. https://doi.org/10.1038/nn2070

Lee, S.-H., Kwan, A. C., Zhang, S., Phoumthipphavong, V., Flannery, J. G., Masmanidis, S. C., … Dan, Y. (2012). Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature*, *488*(7411), 379–383. https://doi.org/10.1038/nature11312

Lefort, S., Tomm, C., Sarria, J., & Petersen, C. C. H. (2009). The Excitatory Neuronal Network of the C2 Barrel Column in Mouse Primary Somatosensory Cortex. *Neuron*, *61*(2), 301–316. https://doi.org/10.1016/j.neuron.2008.12.020

Lennie, P. (2003). The Cost of Cortical Computation. *Current Biology*, *13*, 493–497. https://doi.org/10.1016/S

Li, N., Chen, S., Guo, Z. V, Chen, H., Huo, Y., Inagaki, H. K., … Svoboda, K. (2019). Spatiotemporal constraints on optogenetic inactivation in cortical circuits. *ELife*, *8*, e48622.

Llorca, P. M., Pereira, B., Jardri, R., Brousse, G., & Misdrahi, D. (2016). Hallucinations in schizophrenia and Parkinson's disease: an analysis of sensory modalities involved and the repercussion on patients. *Scientific Reports*, *6*(December), 31852. https://doi.org/10.1038/srep38152

London, M., Roth, A., Beeren, L., Häusser, M., & Latham, P. E. (2010). Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, *466*(7302), 123–127. https://doi.org/10.1038/nature09086

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*, 185–199.

Mardinly, A. R., Oldenburg, I. A., Pégard, N. C., Sridharan, S., Lyall, E. H., Chesnov, K., … Adesnik, H. (2018). Precise multimodal optical control of neural ensemble activity.

Nature Neuroscience. https://doi.org/10.1038/s41593-018-0139-8

Marshel, J. H., Kim, Y. S., Machado, T. A., Quirin, S., Benson, B., Kadmon, J., … Deisseroth, K. (2019). Cortical layer–specific critical dynamics triggering perception. *Science*, *365*, 1–12. https://doi.org/10.1126/science.aaw5202

Massi, B., Donahue, C. H., & Lee, D. (2018). Volatility Facilitates Value Updating in the Prefrontal Cortex. *Neuron*, *99*(3), 598–608. https://doi.org/10.1016/j.neuron.2018.06.033

Mateo, C., Avermann, M., Gentet, L. J., Zhang, F., Deisseroth, K., & Petersen, C. C. H. (2011). In vivo optogenetic stimulation of neocortical excitatory neurons drives brain-state-dependent inhibition. *Current Biology*, *21*(19), 1593–1602. https://doi.org/10.1016/j.cub.2011.08.028

Mayrhofer, J. M., Haiss, F., Helmchen, F., & Weber, B. (2015). Sparse, reliable, and long-term stable representation of periodic whisker deflections in the mouse barrel cortex. *NeuroImage*, *115*, 52–63. https://doi.org/10.1016/j.neuroimage.2015.04.045

McEliece, R., Posner, E. C., Rodemich, E. R., & Venkatesh, S. S. (1987). The Capacity of the Hopfield Associative Memory. *IEEE Transactions on Information Theory*, *33*(4), 461–482.

Mcguire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron*, *84*(4), 870–881. https://doi.org/10.1016/j.neuron.2014.10.013

Meyer, H. S., Egger, R., Guest, J. M., Foerster, R., Reissl, S., & Oberlaender, M. (2013). Cellular organization of cortical barrel columns is whisker-specific. *Proceedings of the National Academy of Sciences*, *110*(47), 19113–19118. https://doi.org/10.1073/pnas.1312691110

Morita, T., Kang, H., Wolfe, J., Jadhav, S. P., & Feldman, D. E. (2011). Psychometric Curve and Behavioral Strategies for Whisker-Based Texture Discrimination in Rats. *PLoS ONE*, *6*(6), e20437. https://doi.org/10.1371/journal.pone.0020437

Murasugi, C. M., Salzman, C. D., & Newsome, W. T. (1993). Microstimulation in visual area MT: effects of varying pulse amplitude and frequency. *The Journal of Neuroscience*, *13*(4), 1719–1729. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8463847

Murphy, B. K., & Miller, K. D. (2009). Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns. *Neuron*, *61*(4), 635–648. https://doi.org/10.1016/j.neuron.2009.02.005

Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., & Churchland, A. K. (2019). Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, *22*(October), 1677–1686. https://doi.org/10.1038/s41593-019-0502-4

Niell, C. M., & Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, *65*(4), 472–479. https://doi.org/10.1016/j.neuron.2010.01.033

O'Connor, D. H., Clack, N. G., Huber, D., Komiyama, T., Myers, E. W., & Svoboda, K. (2010). Vibrissa-based object localization in head-fixed mice. *The Journal of Neuroscience*, *30*(5), 1947–1967. https://doi.org/10.1523/JNEUROSCI.3762-09.2010

O'Connor, D. H., Hires, S. A., Guo, Z. V, Li, N., Yu, J., Sun, Q., … Svoboda, K. (2013). Neural coding during active somatosensation revealed using illusory touch. *Nature Neuroscience*, *16*(7), 958–965. https://doi.org/10.1038/nn.3419

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(June), 607–609.

Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, *14*(4), 481–487. https://doi.org/10.1016/j.conb.2004.07.007

Ölveczky, B. P., Otchy, T. M., Goldberg, J. H., Aronov, D., & Fee, M. S. (2011). Changes in the neural control of a complex motor sequence during learning. *Journal of Neurophysiology*, *106*, 386–397. https://doi.org/10.1152/jn.00018.2011.

Otis, J. M., Namboodiri, V. M. K., Matan, A. M., Voets, E. S., Mohorn, E. P., Kosyk, O., … Stuber, G. D. (2017). Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. *Nature*, *543*(7643), 103–107. https://doi.org/10.1038/nature21376

Ozeki, H., Finn, I. M., Schaffer, E. S., Miller, K. D., & Ferster, D. (2009). Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression. *Neuron*, *62*(4), 578–592. https://doi.org/10.1016/j.neuron.2009.03.028

Pachitariu, M., Stringer, C., & Harris, K. D. (2018). Robustness of Spike Deconvolution for Neuronal Calcium Imaging. *The Journal of Neuroscience*, *38*(37), 7976–7985. https://doi.org/10.1523/JNEUROSCI.3339-17.2018

Pachitariu, M., Stringer, C., Schröder, S., Dipoppa, M., Rossi, L. F., Carandini, M., & Harris, K. D. (2016). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *BioRxiv*. https://doi.org/10.1101/061507

Packer, A. M., Russell, L. E., Dalgleish, H. W. P., & Hausser, M. (2015). Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. *Nature Methods*, *12*(2), 140–146. https://doi.org/10.1038/nmeth.3217

Packer, A. M., & Yuste, R. (2011). Dense, Unspecific Connectivity of Neocortical Parvalbumin-Positive Interneurons: A Canonical Microcircuit for Inhibition? *The Journal of Neuroscience*, *31*(37), 13260–13271. https://doi.org/10.1523/JNEUROSCI.3131-11.2011

Palm, G. (1980). On associative memory. *Biological Cybernetics*, *31*, 19–31.

Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E., & Fellin, T. (2017). Cracking the neural code for sensory perception by combining statistics, intervention and behavior. *Neuron*, *93*(3), 491–507. https://doi.org/10.1111/gcb.13051

Panzeri, S., Ince, R. A. A., Diamond, M. E., & Kayser, C. (2014). Reading spike timing without a clock: intrinsic decoding of spike trains. *Philosophical Transactions of the Royal Society B*, *369*, 1–11.

Pehlevan, C., & Sompolinsky, H. (2014). Selectivity and Sparseness in Randomly Connected Balanced Networks. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089992

Perin, R., Berger, T. K., & Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy*, *108*(13), 5419–5424. https://doi.org/10.1073/pnas.1016051108

Peron, S., Freeman, J., Iyer, V., Guo, C., & Svoboda, K. (2015). A Cellular Resolution Map of Barrel Cortex Activity during Tactile Behavior. *Neuron*, *86*(3), 783–799. https://doi.org/10.1016/j.neuron.2015.03.027

Peron, S., Pancholi, R., Voelcker, B., Wittenbach, J. D., Ólafsdóttir, H. F., Freeman, J., & Svoboda, K. (2020). Recurrent interactions in local cortical circuits. *Nature*, (October 2016), 1–4. https://doi.org/10.1038/s41586-020-2062-x

Petersen, C. C. H. (2007). The functional organization of the barrel cortex. *Neuron*, *56*(2), 339–355. https://doi.org/10.1016/j.neuron.2007.09.017

Plitt, M. H., & Giocomo, L. M. (2019). Experience dependent contextual codes in the hippocampus. *BioRxiv*, 1–30.

Poort, J., Khan, A. G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., … Hofer, S. B. (2015). Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. *Neuron*, *86*(6), 1478–1490. https://doi.org/10.1016/j.neuron.2015.05.037

Rose, D., & Blakemore, C. (1974). Effects of bicuculline on functions of inhibition in visual cortex. *Nature*, *249*, 375–377.

Rudebeck, P. H., & Murray, E. A. (2011). Dissociable Effects of Subtotal Lesions within the Macaque Orbital Prefrontal Cortex on Reward-Guided Behavior. *The Journal of Neuroscience*, *31*(29), 10569–10578. https://doi.org/10.1523/JNEUROSCI.0091-11.2011

Russell, L. E., Yang, Z., Tan, P. L., Fişek, M., Packer, A. M., Dalgleish, H. W. P., … Häusser, M. (2019). The influence of visual cortex on perception is modulated by behavioural state. *BioRxiv*.

Sachidhanandam, S., Sreenivasan, V., Kyriakatos, A., Kremer, Y., & Petersen, C. C. H. (2013). Membrane potential correlates of sensory perception in mouse barrel cortex. *Nature Neuroscience*, (October), 1–9. https://doi.org/10.1038/nn.3532

Sadeh, S., & Clopath, C. (2020). Patterned perturbation of inhibition can reveal the dynamical structure of neural processing. *ELife*, *9*, e52757.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, *346*, 174–176. https://doi.org/10.1016/0021-9797(80)90501-9

Salzman, C. D., Murasugi, C. M., Britten, K. H., & Newsome, W. T. (1992). Microstimulation in Visual Area MT: Effects on Direction Discrimination Performance. *The Journal of Neuroscience*, *12*(6), 2331–2355.

Sanders, H., Wilson, M. A., & Gershman, S. J. (2020). Hippocampal remapping as hidden state inference. *ELife*, *9*, e51140.

Sanzeni, A., Akitake, B., Goldback, H., Leedy, C., Brunel, N., & Histed, M. (2020). Inhibition stabilization is a widespread property of cortical networks. *BioRxiv*.

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2009). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, *11*(1), 53–60. https://doi.org/10.1038/nrn2737

Scholl, B., Pattadkal, J. J., Dilly, G. A., Priebe, N. J., & Zemelman, B. V. (2015). Local Integration Accounts for Weak Selectivity of Mouse Neocortical Parvalbumin Interneurons. *Neuron*, *87*(2), 424–436. https://doi.org/10.1016/j.neuron.2015.06.030

Schölvinck, M. L., Howarth, C., & Attwell, D. (2008). The cortical energy needed for conscious perception. *NeuroImage*, *40*, 1460–1468. https://doi.org/10.1016/j.neuroimage.2008.01.032

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, *122*, 105–123. https://doi.org/10.1016/j.visres.2016.02.002

Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, *4*, 569–579. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7812147

Shemesh, O. A., Tanese, D., Zampini, V., Linghu, C., Piatkevich, K., Ronzitti, E., … Emiliani, V. (2017). Temporally precise single-cell-resolution optogenetics. *Nature Neuroscience*,

20(12), 1796–1806. https://doi.org/10.1038/s41593-017-0018-8

Shuler, M. G., & Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science*, *311*(March), 1606–1610.

Silberberg, G., & Markram, H. (2007). Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells. *Neuron*, *53*, 735–746. https://doi.org/10.1016/j.neuron.2007.02.012

Sippy, T., Lapray, D., Crochet, S., & Petersen, C. C. H. (2015). Cell-Type-Specific Sensorimotor Processing in Striatal Projection Neurons during Goal-Directed Behavior Supplemental Information. *Neuron*, 1–8. https://doi.org/10.1016/j.neuron.2015.08.039

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2018). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653. https://doi.org/10.1038/nn.4650

Steinmetz, N. A., Zatka-haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, *576*(December), 266–273. https://doi.org/10.1038/s41586-019-1787-x

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, *364*(255), 1–11. https://doi.org/10.1126/science.aav7893

Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2020). Cellpose: a generalist algorithm for cellular segmentation. *BioRxiv*, 1–19.

Stüttgen, M., & Schwarz, C. (2008). Psychophysical and neurometric detection performance under stimulus uncertainty. *Nature Neuroscience*, *11*(9), 1091–1099. https://doi.org/10.1038/nn.2162

Tanke, N., Borst, J. G. G., & Houweling, A. R. (2018). Single-cell stimulation in barrel cortex influences psychophysical detection performance. *The Journal of Neuroscience*, *38*(8), 2057–2068. https://doi.org/10.1523/JNEUROSCI.2155-17.2018

Thomson, A. M., & Lamy, C. (2007). Functional maps of neocortical local circuitry. *Frontiers in Neuroscience*, *1*(1), 19–42.

Tolhurt, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, *23*(8), 775–785.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208.

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. I. Orientation and the short-cut. *Journal of Experimental Psychology*, *6*(1), 13–24.

Treiman, D. M. (2001). GABAergic Mechanisms in Epilepsy. *Epilepsia*, *42*(2), 8–12.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007). Schemas and Memory Consolidation. *Science*, *316*(7489), 76–82.

Tsodyks, M. V, Skaggs, W. E., Sejnowski, T. J., & Mcnaughton, B. L. (1997). Paradoxical Effects of External Modulation of Inhibitory Interneurons. *The Journal of Neuroscience*, *17*(11), 4382–4388.

van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science*, *274*(December), 1724–1726.

Walton, M. E., Behrens, T. E. J., Buckley, M. J., Rudebeck, P. H., & Rushworth, M. F. S. (2010). Separable Learning Systems in the Macaque Brain and the Role of Orbitofrontal Cortex in Contingent Learning. *Neuron*, *65*(6), 927–939.

https://doi.org/10.1016/j.neuron.2010.02.027

Waters, J., & Helmchen, F. (2006). Background synaptic activity is sparse in neocortex. *The Journal of Neuroscience*, *26*(32), 8267–8277. https://doi.org/10.1523/JNEUROSCI.2152-06.2006

Watson, B. O., Yuste, R., & Packer, A. M. (2016). PackIO and EphysViewer: software tools for acquisition and analysis of neuroscience data. *BioRxiv*. https://doi.org/https://doi.org/10.1101/054080

Wehr, M., & Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, *426*(November), 860–863.

Whittington, J. C. R., Muller, T. H., Barry, C., & Behrens, T. E. J. (2018). Generalisation of structural knowledge in the Hippocampal-Entorhinal system. *ArXiv*.

Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. J. (2019). The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation. *BioRxiv*.

Wiest, M. C., Thomson, E., Pantoja, J., & Nicolelis, M. A. L. (2020). Changes in S1 Neural Responses During Tactile Discrimination Learning. *Journal of Neurophysiology*, *104*, 300–312. https://doi.org/10.1152/jn.00194.2010.

Wilson, N. R., Runyan, C. A., Wang, F. L., & Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature*, *488*(7411), 343–348. https://doi.org/10.1038/nature11347

Wolf, F., Engelken, R., Puelma-Touzel, M., Weidinger, J. D. F., & Neef, A. (2014). Dynamical models of cortical circuits. *Current Opinion in Neurobiology*, *25*, 228–236. https://doi.org/10.1016/j.conb.2014.01.017

Wolfe, J., Houweling, A. R., & Brecht, M. (2010). Sparse and powerful cortical spikes. *Current Opinion in Neurobiology*, *20*(3), 306–312. https://doi.org/10.1016/j.conb.2010.03.006

Xiong, Q., Znamenskiy, P., & Zador, A. M. (2015). Selective corticostriatal plasticity during acquisition of an auditory discrimination task. *Nature*. https://doi.org/10.1038/nature14225

Ye, Z., Mostajo-Radji, M. A., Brown, J. R., Rouaux, C., Tomassy, G. S., Hensch, T., & Arlotta, P. (2015). Instructing Perisomatic Inhibition by Direct Lineage Reprogramming of Neocortical Projection Neurons. *Neuron*, *88*(3), 475–483. https://doi.org/10.1016/j.neuron.2015.10.006

Yizhar, O., Fenno, L. E., Prigge, M., Schneider, F., Davidson, T. J., Shea, D. J. O., … Deisseroth, K. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*, *477*(7363), 171–178. https://doi.org/10.1038/nature10360

Yoshimura, Y., Dantzker, J. L. M., & Callaway, E. M. (2005). Excitatory cortical neurons from fine-scale functional networks. *Nature*, *5*(February), 2005–2005. https://doi.org/10.1038/nature03291.1.

Zatka-Haas, P., Steinmetz, N. A., Carandini, M., & Harris, K. D. (2020). A perceptual decision requires sensory but not action coding in mouse cortex. *BioRxiv*.

Zhang, Z., Russell, L. E., Packer, A. M., Gauld, O. M., & Häusser, M. (2018). Closed-loop all-optical interrogation of neural circuits in vivo. *Nature Methods*, *15*(December), 1037–1040. https://doi.org/10.1038/s41592-018-0183-z

Ziburkus, J., Cressman, J. R., Barreto, E., & Schiff, S. J. (2006). Interneuron and Pyramidal

Cell Interplay During In Vitro Seizure-Like Events. *Journal of Neurophysiology*, *95*, 3948–3954. https://doi.org/10.1152/jn.01378.2005.

Znamenskiy, P., Kim, M., Muir, D. R., Iacaruso, F., Hofer, S. B., & Mrsic-flogel, T. D. (2018). Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *BioRxiv*.

Znamenskiy, P., & Zador, A. M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature*, *497*(7450), 482–485. https://doi.org/10.1038/nature12077

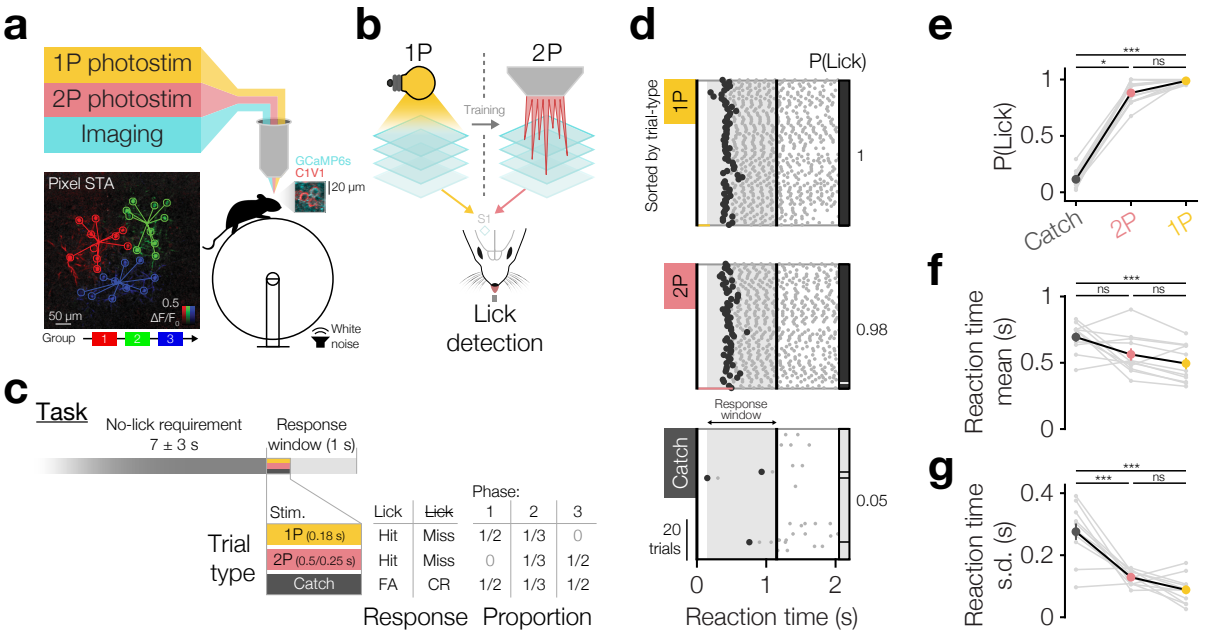

Figure 1 | Driving behaviour with two-photon optogenetics targeted to ensembles of neurons in L2/3 barrel cortex.

a, Schematic of all-optical setup. *Bottom left*: Example of flexible ensemble photostimulation. Three 10 neuron groups in barrel cortex (red, green, blue circles joined by group centroids) were photostimulated sequentially (sequence below Pixel STA). Pixel STA is maximum intensity projection across photostimulus groups of activity in post-photostimulation epoch averaged across trials (*N* = 3 photostimulus groups, 10 trials each). *Inset right*: sub-region of a full imaging FOV in L2/3 barrel cortex expressing GCaMP6s/C1V1-mRuby. **b,** Schematic summarizing the strategy used to train animals to respond to two-photon optogenetic (2P) stimulation. Mice are first trained to respond to one-photon optogenetic (1P) stimulation of barrel cortex (S1) by licking at an electronic lickometer. The power of 1P illumination is reduced until they can be transitioned onto 2P stimulation targeted to specific ensembles of barrel cortex neurons. **c,** Structure of the behavioural task (top) and stimulus probabilities, response type contingencies and training phase structures (bottom). Note that stimulus durations, which vary across stimulus types, are not to scale. **d,** Lick raster from an example Phase 2 behavioural training session during which a mouse received 2P stimulation trials (pink: 200 neurons), catch trials (grey, no stimulus) and 1P stimulation trials (amber: 0.05 mW, untargeted). Trials were delivered pseudo-randomly (see *Methods*) but have been sorted by trial-type for display. All licks shown in grey with first lick highlighted in black. Hits (black) and misses (grey) are indicated as the vertical bar on the right-hand side. Stimulus durations indicated as coloured bars below lick rasters. Behavioural response window indicated as grey shading, label and arrows. **e – g,** Response rates and reaction time mean and standard deviation for different trial types in final Phase 2 session. Animals detect 1P photostimulation and 2P stimulation targeted to 200 neurons to similar extents, at a level far above chance (catch trials), with similar reaction time mean and standard deviation. (*N* = 12 mice, 1 session each). Note only animals which responded on > 2 Catch trials are included for reaction time panels (*f* and *g*) (*N* = 11 animals, 1 session each). All error bars are s.e.m.

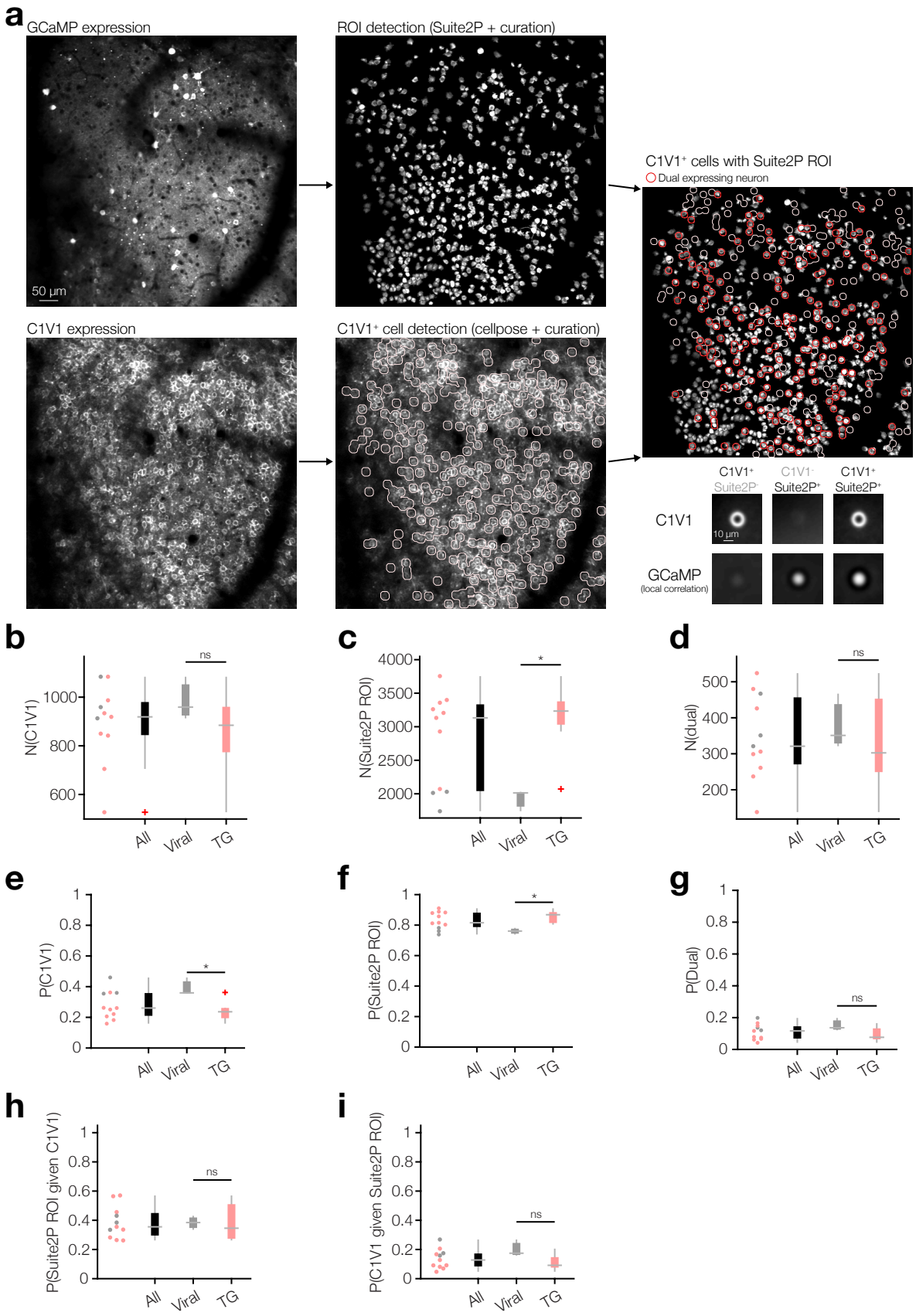

Figure 1 – figure supplement 1 | Indicator and opsin expression overlap analysis.
a, Detection of indicator/opsin expression overlap. Suite2P followed by manual curation was used to detect functional GCaMP expressing neurons from time-series data (*top row*) and Cellpose followed by manual curation was used to detect C1V1 expression from static images (*bottom row*). GCaMP

(*top left*) and C1V1 (*bottom left*) expression in an example plane from a volumetric FOV with corresponding Suite2P ROIs (*top middle*) and detected C1V1^{+} neurons (*bottom middle*). C1V1^{+} neurons with a Suite2P ROI are circled in red (*right*; see *Methods*) and the average expression pattern in both C1V1 expression and GCaMP local correlation images is shown for each of the 3 possible expression configurations (*bottom right insets*). *N* = 6250 C1V1^{+}/Suite2P^{-} neurons, 27084 C1V1^{-}/Suite2P^{+} neurons and 3810 C1V1^{+}/Suite2P^{+} neurons across 11 volumetric FOVs from 6 animals, 1 – 2 volumetric FOVs each. **b,** Number of C1V1^{+} neurons in each volumetric FOV in all mice (black), dual injected WT mice (grey) and opsin injected transgenic GCaMP mice (pink). See below and *Methods* for details. **c,** Number of Suite2P ROIs in each volumetric FOV. **d,** Number of dual expressing neurons in each volumetric FOV (C1V1^{+} neurons with Suite2P ROI). **e,** Proportion of C1V1^{+} neurons out of total number of neurons that have a Suite2P ROI and/or are C1V1^{+}. **f,** Proportion of neurons with a Suite2P ROI out of total number of neurons (same denominator as *e*). **g,** Proportion of dual expressing neurons (C1V1^{+} with Suite2P ROI) out of total number of neurons (same denominator as *e*). **h,** Proportion of C1V1^{+} neurons with a Suite2P ROI. **i,** Proportion of Suite2P ROIs that are C1V1^{+}. For all group data panels *N* = 11 volumetric FOVs from the 2P psychometric curve sessions (**Figure 2 – 4**), 3 FOVs from WT (C57/BL6) mice co-injected with hSyn-GCaMP6s/CaMKII-C1V1-Kv2.1 (grey in all plots) and 8 FOVs from transgenic (Emx1Cre;CaMKIIa-tTA;Ai94) mice injected with CaMKII-C1V1-Kv2.1 (pink in all plots). Aggregate data across expression strategies are black in all plots.

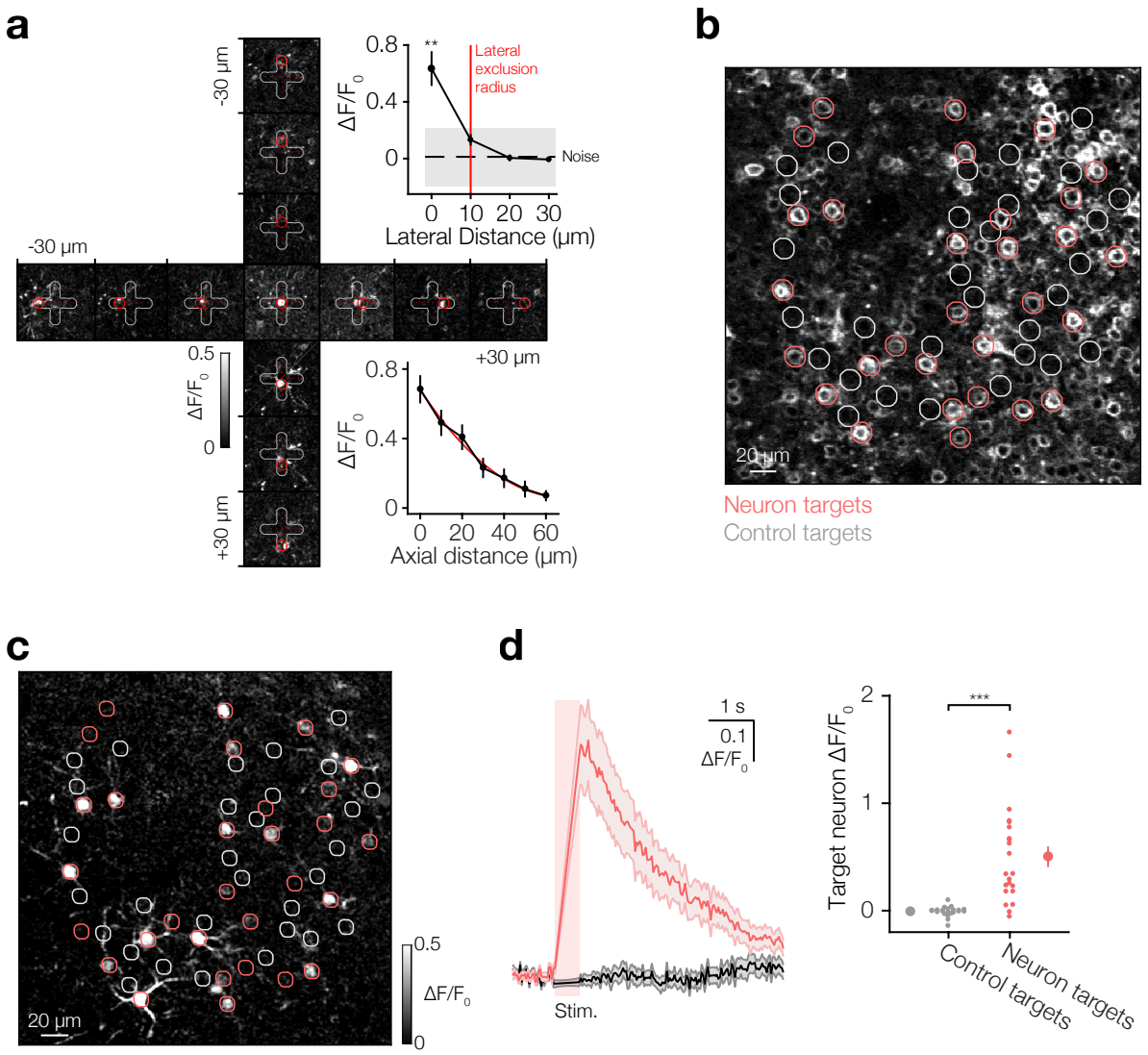

Figure 1 – figure supplement 2 | Spatial resolution of spiral-scanned two-photon optogenetic activation of Kv2.1-C1V1.

a, *Left*: A cross-shaped grid of photostimulation targets was centred on a single target neuron and each photostimulation target was activated 10 times in random order (10 µm inter-target distance, 30 µm maximum distance from centre, 10 s inter-trial interval, 10 x 20 ms spirals at 20 Hz, 6 mW/neuron). *Top right inset*: Lateral (*xy*) resolution of photostimulation of the central target neuron. Only the target position centred over the target neuron result in significant activation. Lateral HWHM = 5 µm (2-degree polynomial fit). *N* = 1 neuron, 10 trials per stimulation site. Note the lateral exclusion radius (10 µm) used to define target neurons in subsequent analyses is plotted in red. All neurons within this region were considered potential target neurons (see *Methods*). *Bottom right inset*: Axial (*z*) resolution of photostimulation. Photostimulation targets were axially translated from 60 µm above target neurons down onto their cell-body locations in the objective focal plane in 10 µm steps using SLM-based axial displacement while the target neuron responses were consistently imaged at the objective focal plane (separate experiments from lateral resolution). HWHM = 20 µm (2-degree polynomial fit). *N* = 20 neurons, 10 trials per axial offset. To account for these considerable axial off-target effects we defined all neurons within the lateral exclusion zone of 10 µm (see above) to be potential target neurons irrespective of their axial displacement. **b,** Example Kv2.1-C1V1-mScarlet expression image with neuron (pink) and control (grey) targets overlaid as circles. **c,** Maximum intensity projection across two pixel-wise STAs triggered off neuron and control site stimulation. Neuron stimulation results in robust activation of targeted neurons, control stimulation results in no activation. *N* = 10 trials. **d,** *Left*: Extracted traces from target neuron ROIs during both neuron site stimulation (pink) and control site stimulation (grey). *Right*: post-stimulus $\Delta F/F_0$ from target neuron ROIs. Transients evoked in target

neuron ROIs during neuron site stimulation are completely absent during stimulation of adjacent control sites, resulting in no recorded $\Delta F/F_0$ response. N = 20 neurons, 10 trials each. All error bars and shading are s.e.m.

a Typical training schedule (~21 days post injection):

Session	Reward	Trial-type order	Task	Phase
0	Passive	-	Habituation	1
1	Passive	-	1P 10 mW, Catch	1
2	Active	Sequential	1P 10, 7.5, 5 mW, Catch	1
3 - 4	Active	Sequential	1P 5, 2.5, 1.75, 0.5, 0.25 mW, Catch	1
5 - 6	Active	Interleaved	1P 250, 200, 150, 100, 50 µW, Catch	1
7 - 8	Active	Interleaved	1P 100, 80, 60, 40, 20 µW, Catch	1
9	Active	Interleaved	1P 50 µW, 2P 200 cells, Catch	2
10 - 11	Active	Interleaved	1P 50 µW, 2P 200/100 cells, Catch	2
12 - n	Active	Interleaved	2P <= 200 cells, Catch	3

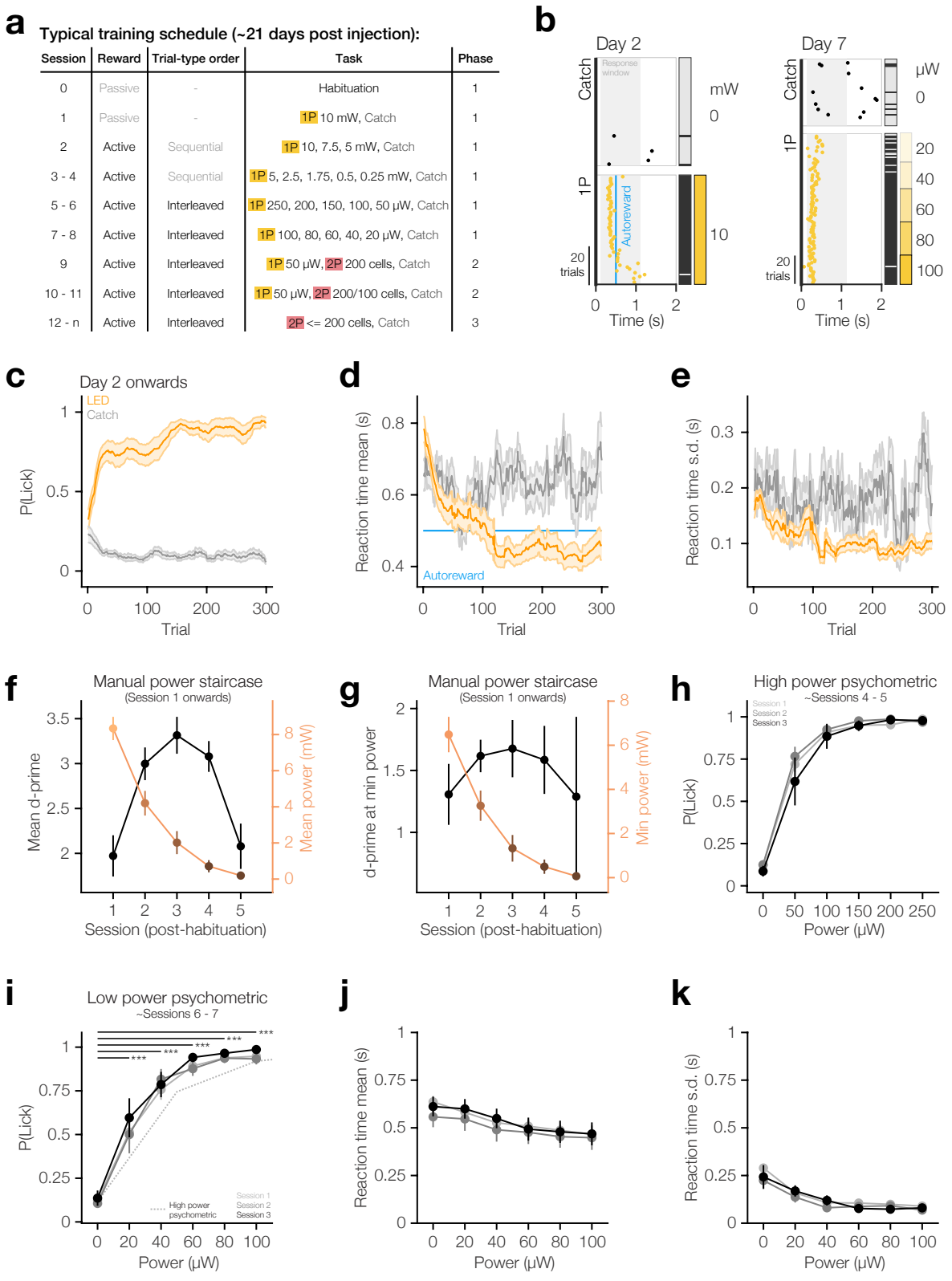

Figure 1 – figure supplement 3 | 1P training protocol.
a, Typical training protocol to transition a naïve mouse to being able to detect two-photon stimulation via a series of one-photon priming phases. Note that exact duration of each component varied by mouse (see **Figure 1 – figure supplement 5f**). **b,** Lick rasters from an example first training session (*left*) and an example low power psychometric curve session (*right*) for a habituated mouse showing first licks on each trial (circles), auto-reward delivery time (blue vertical line) and analysis window (grey shading). Trials were delivered pseudorandomly and are sorted for display only. Hits (black) and misses

(grey) are indicated as the vertical bar to the right of the raster, followed by coloured boxes indicating LED power on the far right. **c – e,** Moving average response rate (*c*), average reaction time (*d*) and standard deviation of reaction time (*e*) for the first 300 LED (amber) and catch (grey) trials using a 20 trial window. **f,** Mean power (*right*) and mean d-prime (*left*) over initial manual power staircase sessions. **g,** Minimum power used in each session (*right*) and d-prime at that power (*left*) over initial manual power staircase sessions. **h,** Response rates on three high power psychometric curve sessions. *N* = 22 animals that completed at least one high power psychometric curve session (other animals were transitioned straight to low power psychometric curves). **i – k,** Response rates (*i*), mean reaction time (*j*) and standard deviation of reaction time (*k*) on three low power psychometric curve sessions. *N* = 22 animals that completed at least one low power psychometric curve session (other animals were transitioned straight to subsequent training phases). All error bars are s.e.m. For all group data panels *N* = 26 mice unless otherwise stated, however some data-points may have fewer *N* as not all animals completed the same number of sessions (see **Figure 1 – figure supplement 5f** for summary).

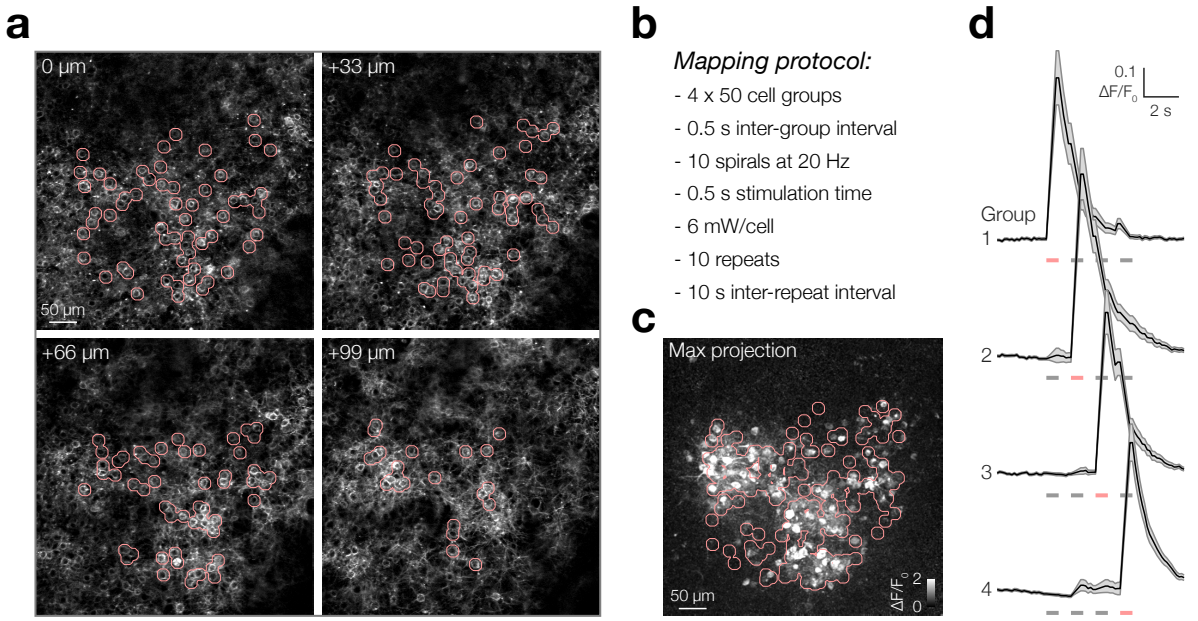

Figure 1 – figure supplement 4 | Example pre-training selection and mapping of 200 neurons in L2/3 barrel cortex.

a, Kv2.1-C1V1-mScarlet expression from an example four-plane imaging volume with stimulation targets overlaid in pink. Neurons are chosen arbitrarily, with the only criteria being that neurons are expressing C1V1 and not too far from FOV centre. Depths relative to most superficial plane (~130 µm below pia). **b,** Pre-training mapping protocol used to assess photostimulatability. Note 200 neurons are mapped in 4 x 50 neuron groups. **c,** Example pixel-wise STA maximum intensity projected across the 4 stimulation groups and the 4 planes with targets overlaid in pink. **d,** Trial-averaged STA traces from ROIs in (c), 50 neurons per group. Note the temporal offset (0.5 s) between activation of each group. Shading denotes s.e.m. *N* = 1 FOV, 1 animal, 200 neurons, 10 trials.

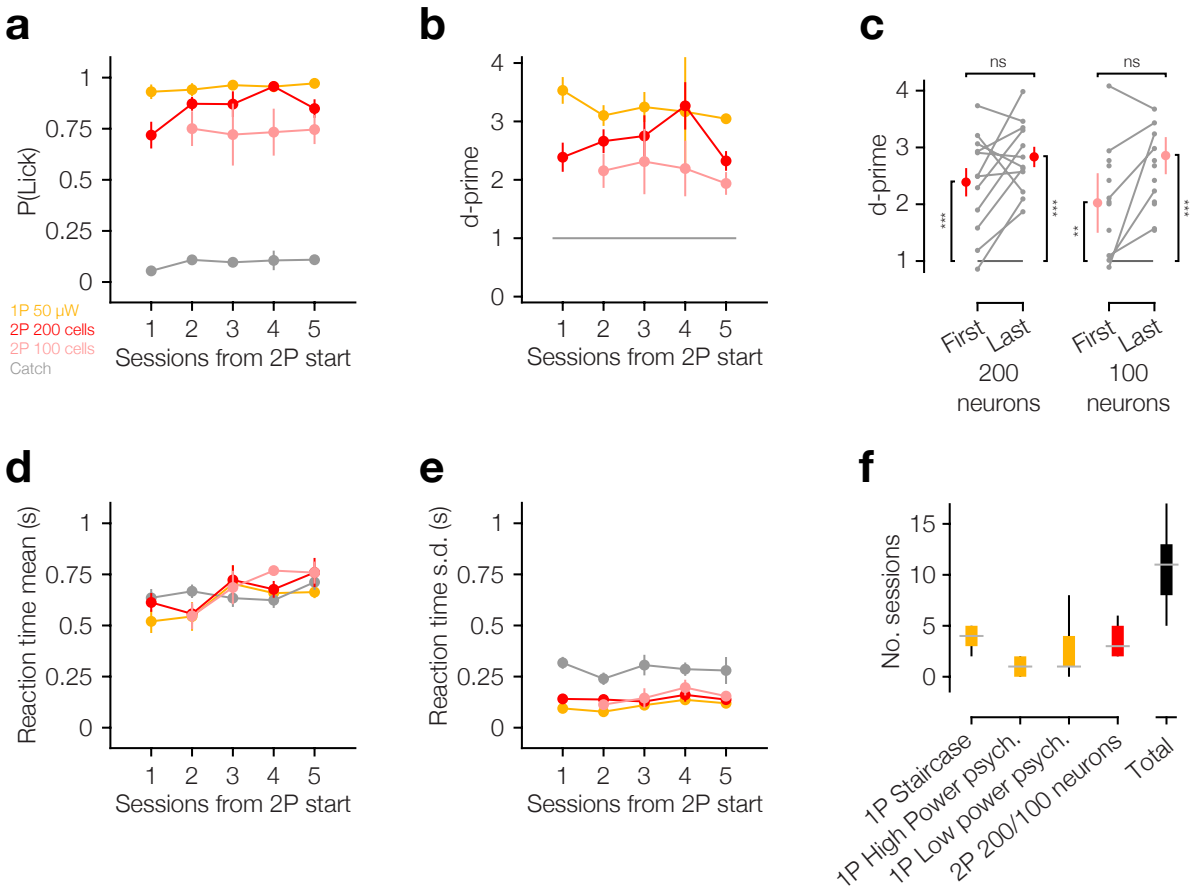

Figure 1 – figure supplement 5 | Rapid transfer learning from 1P to 2P optogenetic stimuli.
a – b Response rates for different 1P and 2P stimulus types from the first session using 2P stimuli, quantified as proportion of trials where animal licked (*a*) and d-prime relative to catch trials (*b*). Note that this includes a combination of Phase 2 (LED/2P/catch trials) and Phase 3 (2P/catch trials) sessions. **c,** Quantification of d-prime relative to our learning criterion (d-prime > 1) for the first and last session using 2P stimuli of different magnitudes. Note that some animals only had 1 session using 2P stimulation of 100 neurons. This session is duplicated in the *First* and *Last* tests against d-prime > 1, but not included in the test for improvement over time. As such these data points appear as grey dots that are not joined by horizontal grey lines. **d – e,** Reaction time mean (*d*) and standard deviation (*e*) for different trial types from the first session using 2P stimuli. Trial types with < 3 responses were excluded from reaction time analyses. **f,** Number of sessions of each step in the training paradigm (coloured bars) and total number of sessions from beginning of 1P training to end of initial 2P training (black bar). For all plots maximum $N = 12$ animals, though individual data-points may have less as not all animals had the same number of training sessions.

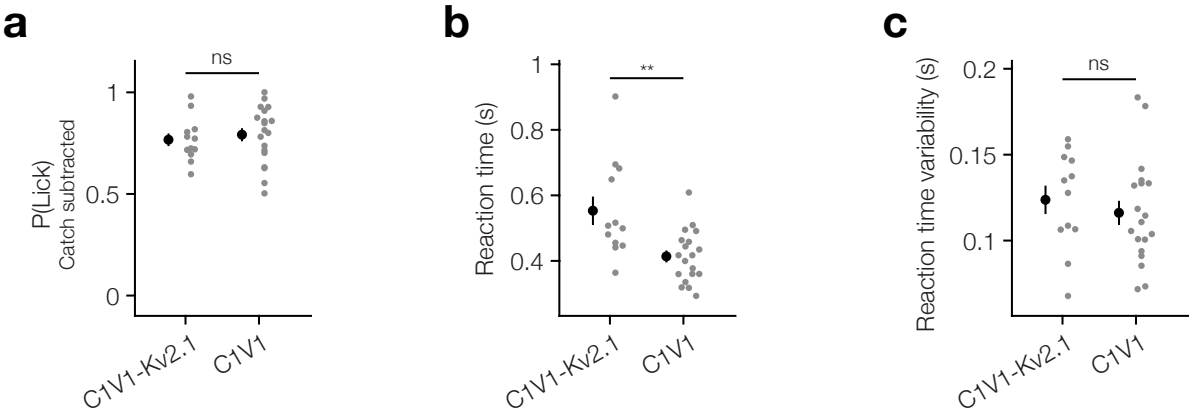

Figure 1 – figure supplement 6 | Comparison of behavioural response to somatic and non-somatic C1V1.

a, Animals detect stimulation of somatic and non-somatic C1V1 to similar extents. **b,** Animals response slightly quicker to stimulation of non-somatic C1V1. **c,** Reaction times are similarly consistent across somatic and non-somatic C1V1. All error bars are s.e.m. *N* = 12 C1V1-Kv2.1 mice, 19 C1V1 mice.

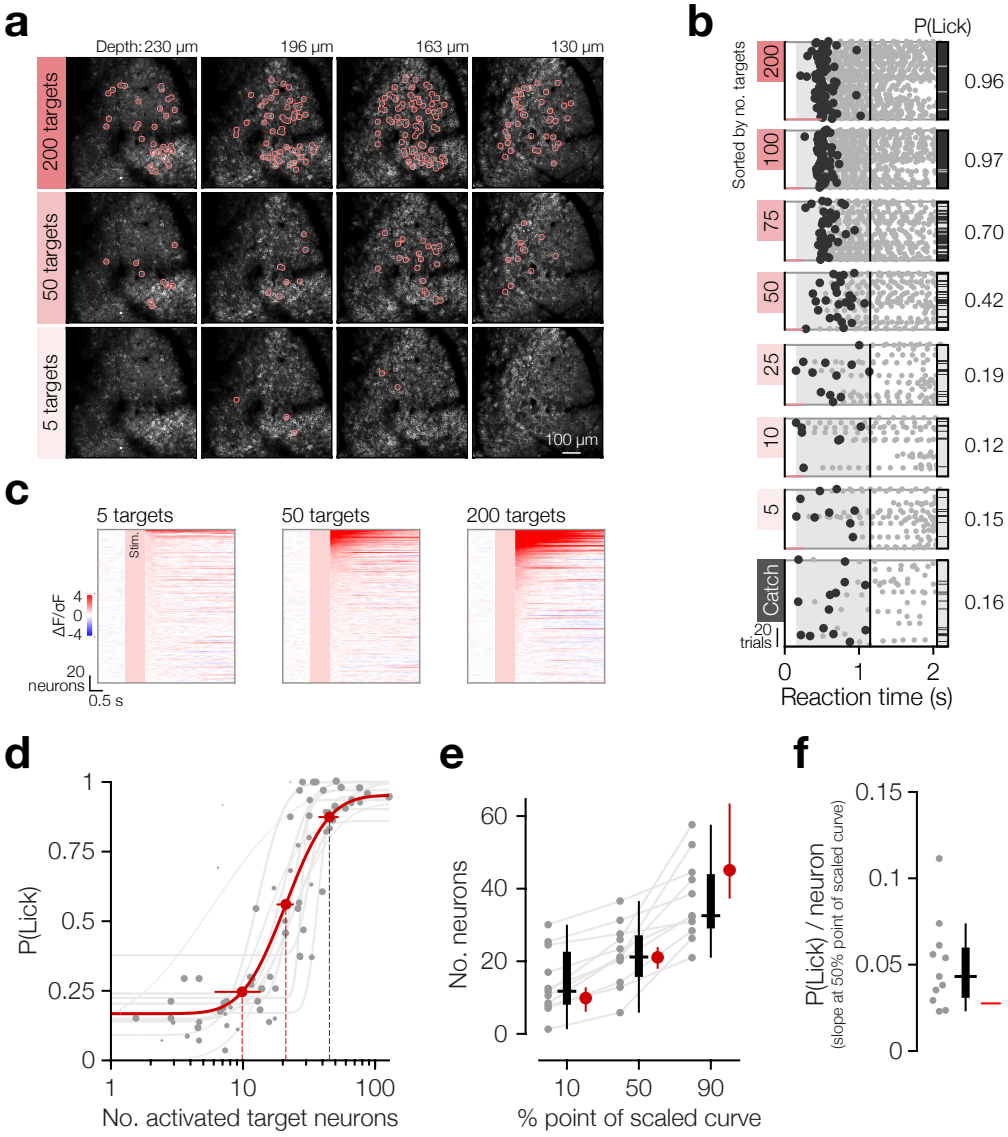

Figure 2 | Animals detect the targeted activation of tens of neurons.
a, Example imaging volumes from an experiment showing 200 (top), 50 (middle) and 5 (bottom) targeted C1V1-expressing neurons. **b,** Example lick raster concatenating an animal's two psychometric curve testing sessions (note data in panels *a* and *c* correspond to the first of these sessions). Trials were delivered pseudo-randomly (see *Methods*) but have been sorted by trial-type for display. Stimulus durations are indicated by coloured bars along the bottom of each raster. Animals respond on more trials and with less variable timing as more neurons are targeted. **c,** Example responses across the top 200 most responsive neurons in the 200 target zones (see *Methods*, **Figure 2 – figure supplement 1**). Neurons have been sorted separately in each plot. Pink boxes indicate the stimulus artefact exclusion epoch which is consistent across all trial-types (see *Methods* for definition). **d,** The psychometric function relating the number of activated target neurons to the behavioural detection rate for all 2P psychometric curve sessions. Individual data (grey dots) are grouped by trial type within session (number of target zones) and plotted as the average number of target neurons activated across all trials of each type. Data point size indicates the number of trials of each type (29 ± 8 trials, range 11 – 44, across data points). Individual psychometric curve fits for each session are plotted (grey lines) weighted by the total number of stimulus trials in the session (202 ± 50 trials, range 97 – 245, across sessions). The aggregate psychometric curve fit across all trial types, all sessions, is plotted in red with the 10%, 50% and 90% points of the scaled curve indicated ± confidence intervals. The number of activated target neurons required to reach these points are indicated by the red dashed vertical lines. Note that individual curves are often steeper than the aggregate curve. **e,** The number of neurons required to reach the 10%, 50% and 90% points of the scaled psychometric function in *d*.

Grey data points/lines are quantified from individual psychometric curve fits (grey lines in *d*) and summarized by the black boxplot distributions. Red data points are quantified from the aggregate psychometric curve fit (red line/dots in *d*) ± confidence intervals. **f,** The slope at the 50% point of the scaled curve corresponding to the additional probability of detection (P(Lick)) added per target neuron activated. Grey data points are quantified from individual psychometric curve fits (grey lines in *d*) and are summarized by the black boxplot distributions. The red line is quantified from the aggregate psychometric curve fit (red line in *d*) for which no confidence intervals can be calculated (see *Methods*). *N* = 11 sessions, 6 mice, 1 – 2 sessions each. All fit parameter error bars are confidence intervals, all boxplot quantification is median with 25^{th}/75^{th} percentile box and whiskers extending to the last data points not considered outliers (see *Methods*).

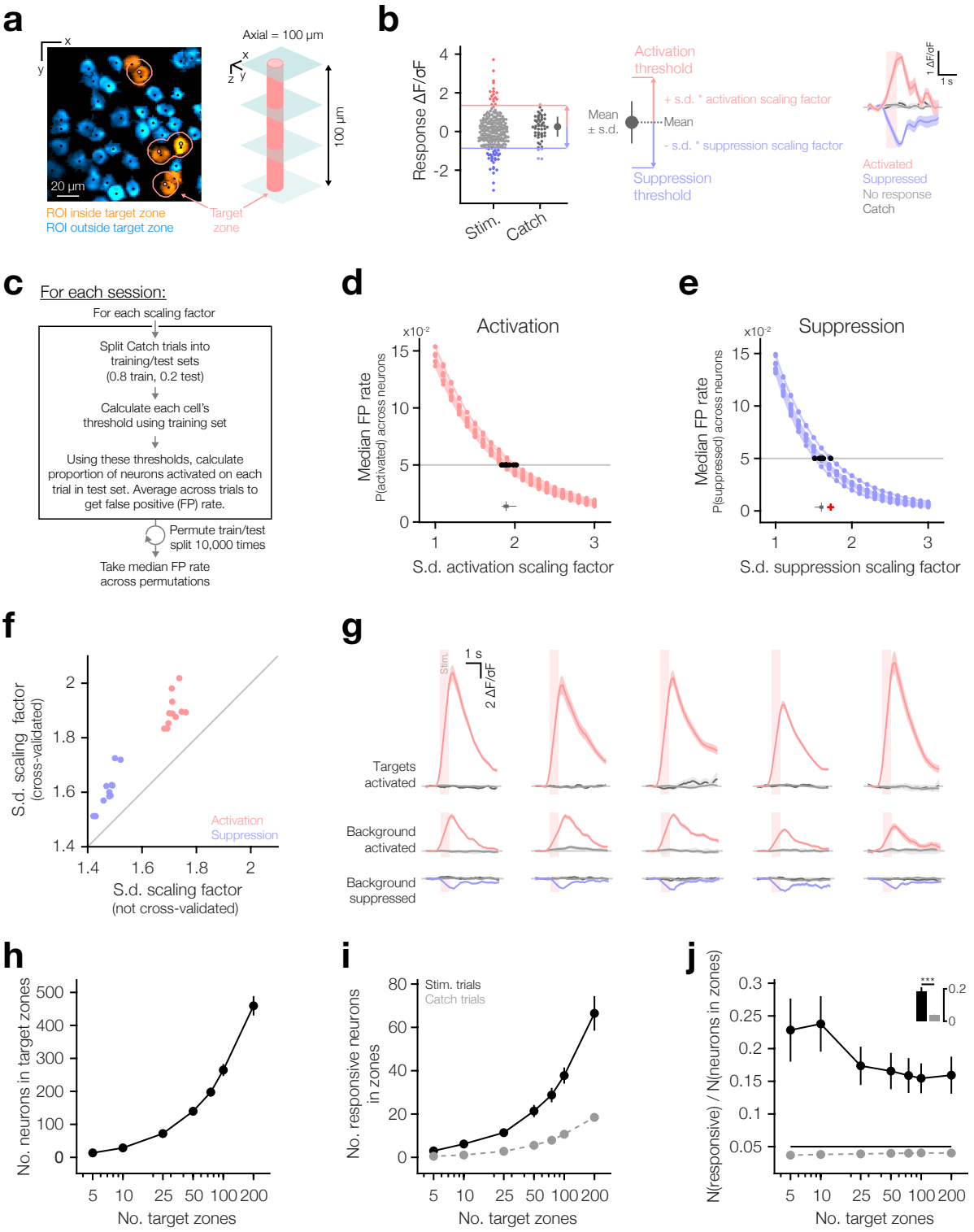

Figure 2 – figure supplement 1 | Quantification of neuronal responses.
a, Example target zones. *Left:* section of a plane in a volumetric FOV showing Suite2P ROIs (coloured regions), Suite2P centroids (black dots), target co-ordinates (pink dots) and lateral extent of target zone (pink circles; 10 µm radius). ROIs with centroids within the target exclusion zones are considered potential targets (orange ROIs), ROIs outside these regions are considered background neurons (blue ROIs). *Right:* schematic illustrating axial extent of target zone. **b,** Rationale behind activation and suppression thresholds calculated from correct reject (CR) catch trial responses and illustration of the results of our threshold procedure (described in *c – e*). *Left*: trial-wise responses for a single non-target neuron on stimulus trials of all types and CR catch trials. The mean and standard deviation of

responses on CR catch trials can be used to infer activation and suppression thresholds that separate stimulus evoked activation (red) and suppression (blue) responses from the majority of CR catch trial responses. *Middle:* activation and suppression thresholds are computed as the mean + or – the standard deviation respectively, scaled by a scaling factor (separate scaling factors for activation and suppression). *Right*: average responses on trials where this neuron was activated (red), suppressed (blue) or unresponsive (light grey). Catch trials are also included (dark grey). Such a procedure separates positive and negative going responses from non-responsive trials which should themselves be indistinguishable from catch trials. **c,** Procedure for inferring cross-validated positive and negative standard deviation scaling factors that yield a 5% false positive response rate on catch trials for each session's volumetric FOV. For each session's data, and each threshold type (i.e. activation threshold), we sweep through a series of scaling factors. For each factor we permute a 80:20 train:test split of correct reject (CR) catch trials 10,000 times. On each permutation we use the scaling factor and that permutation's training CR catch trials to calculate each neuron's threshold. Using these neuron-wise thresholds we then compute the proportion of neurons activated on each testing CR catch trial and average across trials (false positive rate; FP). We then take the median FP across all permutations at this scaling factor before moving to the next scaling factor. We can then infer the scaling factor that yields a 5% FP rate on testing CR catch trials across permutations. **d,** Relationship between standard deviation scaling factor and FP rate for the activation threshold. Data points are empirically quantified values, fits are cubic interpolations. The optimal scaling factor for each experiment (horizontal boxplot) is inferred by finding the scaling factor for which the fitted FP rate is 5% (horizontal line). **e,** Relationship between the standard deviation scaling factor and FP rate for the suppression threshold. Data conventions and scaling factor inference the same as *d*. **f,** Scaling factors that yield 5% false positive rate when all correct reject catch trials are used (not cross-validated, i.e. thresholds inferred and tested on same trials) compared to those yielded by the cross-validation procedure described in *c – e*. Cross-validation yields more stringent thresholds. **g,** Example traces from one session for 5 activated target neurons (top row), 5 activated background neurons (middle row) and 5 suppressed background neurons (bottom row) illustrating responses yielded by our procedure. Averages across all activated/suppressed trials are shown in red/blue and averages across catch trials/non-responsive stimulus trials are shown in light and dark grey respectively. Note that neurons in each column are unrelated. **h,** The number of neurons in all target zones for each trial type (as defined in *a* and *Methods*). **i,** The number of responsive neurons in all target zones for each trial type on both stimulus trials (black) and catch trials (grey). **j,** Proportion of responsive neurons in all target zones for each trial type on both stimulus trials (black) and catch trials (grey) quantified as the number of responsive neurons in target zones (*i*) divided by the number of neurons in target zones (*h*). *Inset:* Proportion of target zone neurons that are responsive on stimulus trials compared to catch trials. N = 11 sessions, 6 animals, 1 – 2 sessions each for all group data plots. Error bars and shading are mean ± s.e.m.

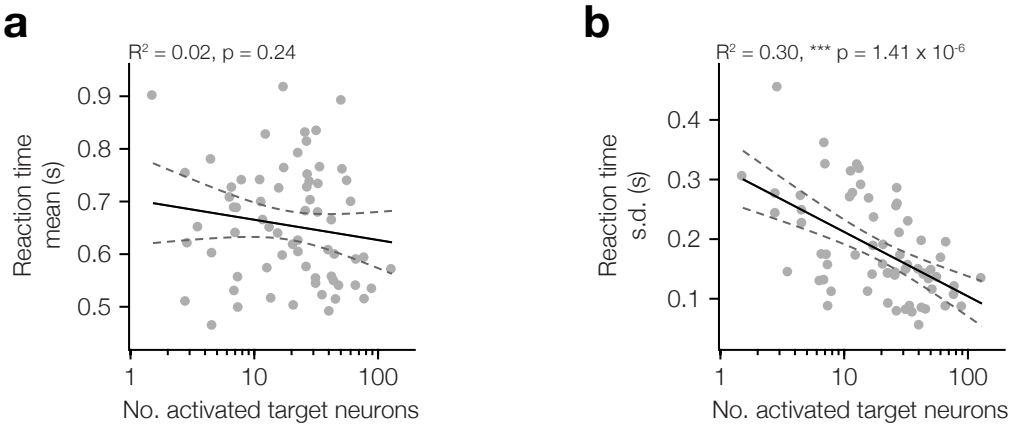

Figure 2 – figure supplement 2 | Reaction time standard deviation, but not mean, scales with the number of target neurons activated.

a – b, Relationship between mean (*a*) and standard deviation (*b*) of reaction time and the number of activated target neurons.

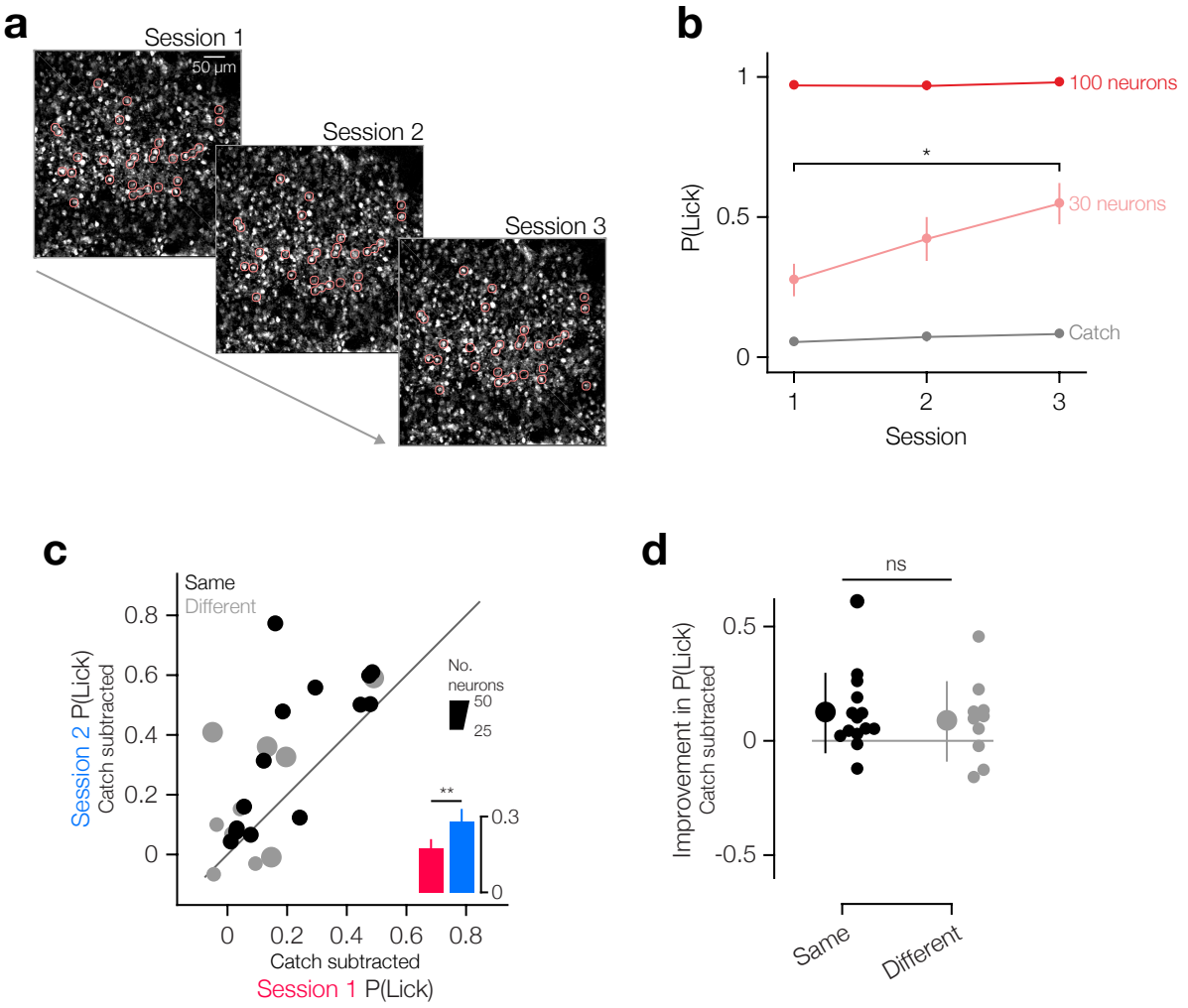

Figure 2 – figure supplement 3 | Detection of small ensembles of neurons improves across days irrespective of whether the same neurons were targeted.
a, FOV from training sessions across 3 consecutive days where we stimulated the same 30 neurons each day **b,** Average behavioural response rates for 100 neuron, 30 neuron and Catch trial conditions for all sessions where we stimulated the same neurons across days (*N* = 14 mice). **c,** Improvement in response rates from session 1 to session 2. Sessions in which we targeted the same neurons are shown in black and sessions where we targeted different neurons in grey. Data-point size denotes number of targeted neurons, scale inset (same neurons group *N* = 14 animals, different neurons group *N* = 5 animals, 2 stim types each). **d,** Sessions targeting the same neurons showed a similar amount of improvement as sessions targeting different neurons (same neurons group *N* = 14 animals, different neurons group *N* = 5 animals, 2 stim types each). All error bars are s.e.m.

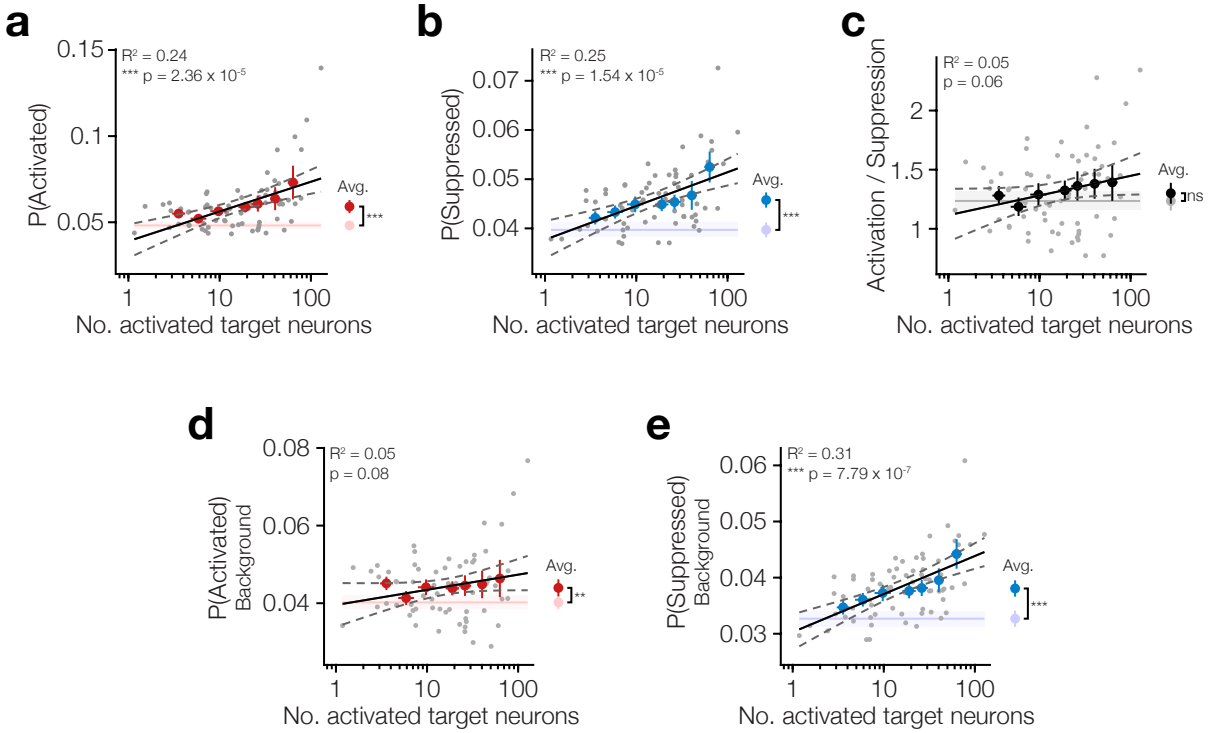

Figure 3 | Increasing target activation is matched by background network suppression.
a, The proportion of neurons activated across all neurons (targets and background) increases as more target neurons are activated. *Inset right:* average activation across all trial types is increased on stimulus trials compared to catch. **b,** The proportion of neurons suppressed across all neurons (targets and background) increases as more target neurons are activated. *Inset right:* average suppression across all trial types is increased on stimulus trials compared to catch. **c,** The ratio of activation and suppression is similar to that observed on catch trials (*inset right*) and isn't strongly modulated by the number of activated target neurons. **d,** Stimulation of target neurons causes mild activation of background neurons (targets excluded; *inset right*) but this is not modulated by the number of target neurons activated. **e,** Stimulation of target neurons causes suppression of background neurons (targets excluded; *inset right*) which increases as more target neurons are activated. All data are hit:miss matched to remove potential lick signals (see **Figure 3 – figure supplement 1** and *Methods*). For all plots $N = 11$ sessions, 6 animals, 1 – 2 sessions each. Some trial types from some sessions are excluded for having too few hits or misses to be able to match the hit:miss ratio. Error bars and shading are mean ± s.e.m; data points, error bars and linear fits are stimulus trials, shading is catch trials; grey data points: individual trial types, individual sessions; coloured error bars: data averaged within trial type (number of target zones) across sessions; linear fits are to individual data points; fits are reported ± 95% confidence intervals.

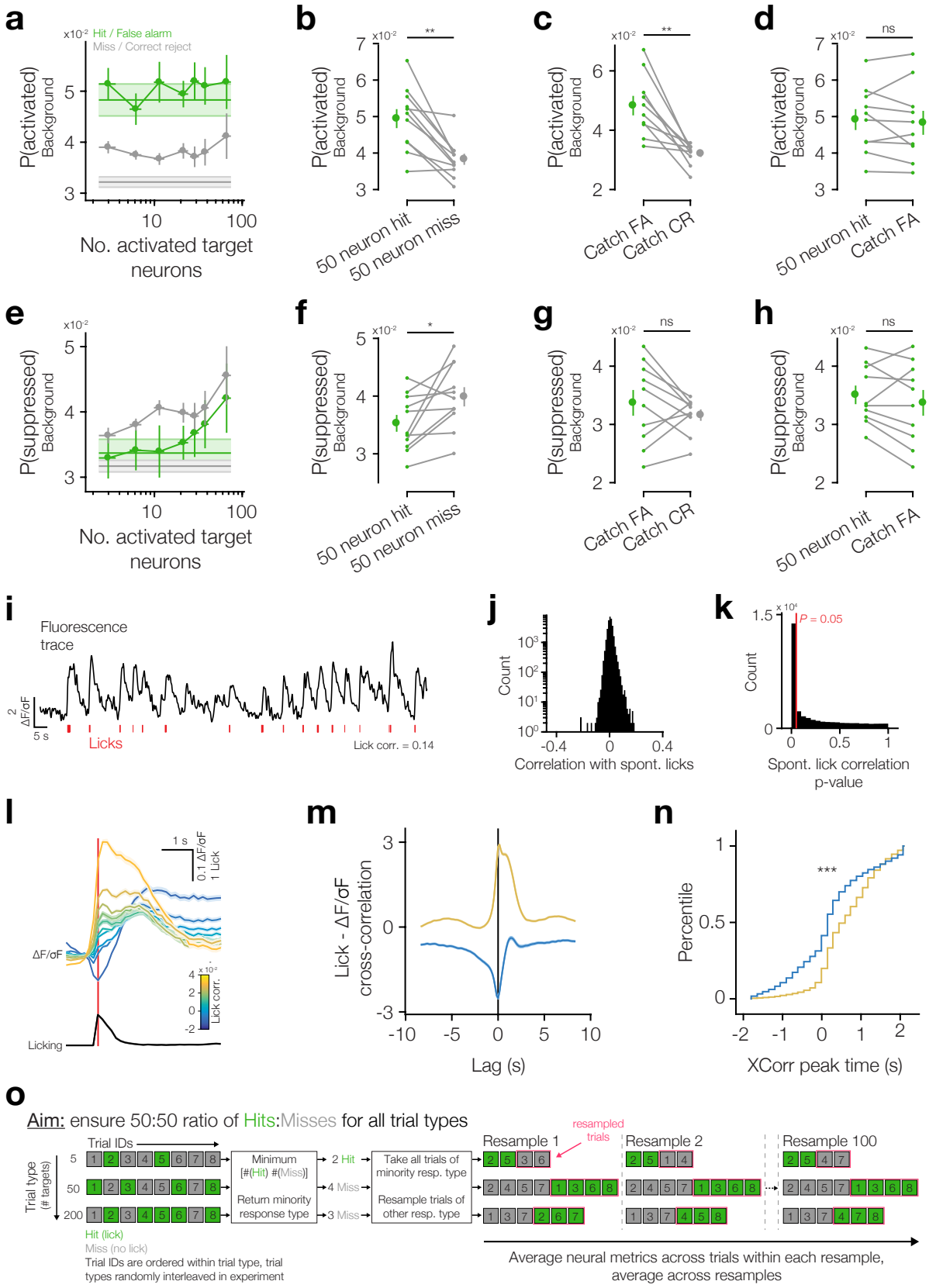

Figure 3 – figure supplement 1 | Comparison of network activity on Hits and Misses for both threshold Go trials and Catch trials in an effort to quantify and account for lick responses.
a, The proportion of background neurons activated when different numbers of target neurons are activated on hits/false alarms (green) and misses/correct rejects (grey) on stimulus trials (error bars) and catch trials (shading). Background neurons become more active on stimulus hits and catch false

alarms and there does not seem to be much modulation of the amount of background activation by the number of activated target neurons. Note that the background neuron activation response on stimulus trials overlaps with the response on catch trials across the full range of number of target neurons activated. **b,** The proportion of activated background neurons is higher on stimulus hit trials than miss trials. Note that we used the 50 target zone stimulation trial type as it has a roughly equal number of hits and misses. **c,** The proportion of activated background neurons is greater on catch false alarms (FA; when animals lick) then on catch correct rejects (CR; when the animal don't lick). **d,** The proportion of activated background neurons does not differ between 50 target zone stimulus trial hits and catch trial false alarms raising the concern that this response is due to licking, not photostimulation. **e,** The proportion of background neurons suppressed when different numbers of target neurons are activated. Same colour conventions as in *a*. Stimulus hits are associated with decreased levels of background suppression than misses, and this appears to be somewhat modulated by the number of activated target neurons. Catch trial false alarms are associated with marginally more suppression than catch trial correct rejections. **f,** The proportion of suppressed background neurons is lower on hits than misses. **g,** The proportion of suppressed background neurons does not differ significantly between false alarms and correct rejects on catch trials. **h,** The proportion of suppressed background neurons does not differ between 50 target zone stimulus trial hits and catch trial false alarms. **i,** Fluorescence trace from an example neuron with photostimulation trial epochs removed (black trace) with licks denoted below (red ticks). This neuron had a lick correlation of 0.14. The fluorescence trace has been Gaussian smoothed (σ = 1s) for display only. **j,** Distribution correlation values across all neurons recorded. **k,** Distribution of *P*-values for correlation values reported in *b*. 46 ± 11 % of neurons are lick modulated (α = 0.05). **l,** *Top:* Average traces from neurons binned into 10 equal bins by their lick correlation (trace colours) aligned to the onset of spontaneous licking (red line). *Bottom:* average across all spontaneous lick bouts. **m,** Lick – fluorescence cross-correlation for all lick responsive neurons positively modulated (yellow) or negatively modulated (blue) by licking. **n,** Distribution of the time of maximum |cross-correlation| for positively modulated (yellow) and negatively modulated (blue) neurons in a 4 s window centred on the 0 time-lag of the cross-correlation. $P = 2.75 \times 10^{-160}$ Mann Whitney U-Test, N = 9547 positively lick-modulated neurons and 4365 negatively modulated neurons. **o,** Schematic illustrating the procedure used to ensure a 50:50 ratio of hits:misses for each trial type. This is done to remove the effect of increased recruitment of licking, and thus lick evoked neural responses, by stimulating target ensembles of increasing size. For a given trial type, say 5 target zones, we find the number of hits and misses and find the minority response type. For 5 target zone trials, which aren't reliably detected by animals, this will likely be hits. We then take all hit trials and combine them with random resamples of miss trials of the same number as hit trials. For example if we only have 2 hits and 6 misses then for each random resample we would take all 2 hits and 2 random misses. In this way we ensure that we have an equal number of hit and miss trials. We then calculate neuron response metrics across the subset of hits and misses in each resample and average these metrics across resamples. All shading is s.e.m. N = 11 sessions, 6 mice, 1 – 2 sessions each.

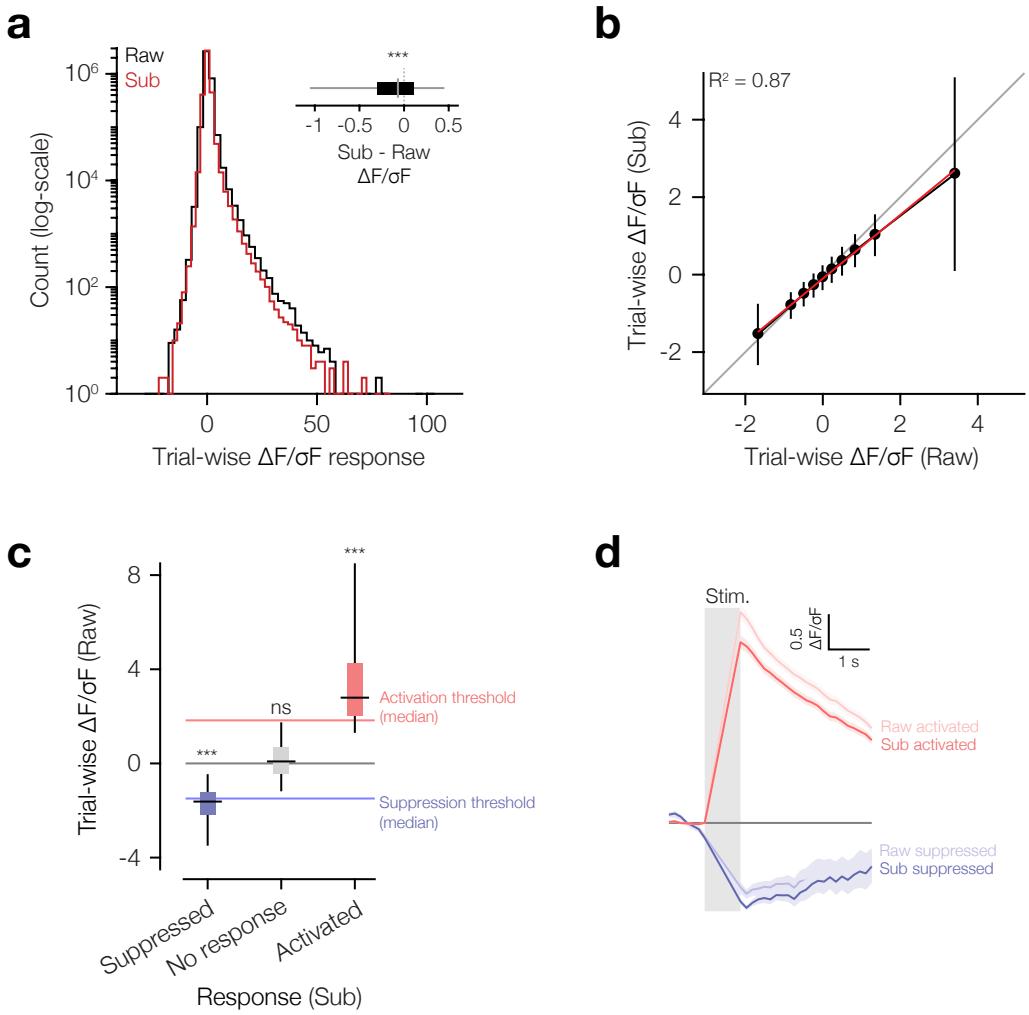

Figure 3 – figure supplement 2 | Neuropil subtraction has a small effect on response amplitude but it is not the sole cause of negative going responses.

a, Distribution of trial-wise responses on 50+ neuron stimulation trials calculated from raw (Raw; black) and neuropil subtracted (Sub; red) traces (N = 30,894 neurons, 1312 trials). A large fraction of negative responses are observed even in raw traces without neuropil subtraction. *Inset*: Difference between each neuron's trial-wise response when responses were calculated from neuropil subtracted or raw traces (Sub – Raw). Responses are slightly reduced when traces are neuropil subtracted (-0.16 ± 0.58 ΔF/σF, P = 0 Wilcoxon signed-rank test). **b,** Neuropil-subtracted trial-wise responses are highly correlated with raw trial-wise responses (R^2 = 0.87, P = 0) though slightly reduced (β_0 = -0.10 ± 5.32 x 10^{-4}, P = 0) with this reduction mainly limited to large positive going responses (β_1 = 0.82 x 10^{-1} ± 3.31 x 10^{-4}, P = 0). Red: linear fit to data, Black: neuropil-subtracted data binned into deciles by raw trial-wise response and averaged within decile. Data are mean ± s.d. **c,** Trial-wise responses classified as suppressed, activated and no response using neuropil-subtracted traces also show negative (-1.75 ± 1.05 ΔF/σF, P = 0 Wilcoxon signed-rank test), roughly zero (0.16 ± 0.94 ΔF/σF, P = 0, Wilcoxon signed-rank test) and positive (3.72 ± 3.18 ΔF/σF, P = 0 Wilcoxon signed-rank test) trial-wise responses calculated from raw traces in the same trials. These groups also differed significantly from each other (P = 0 Kruskal-Wallis test; Suppressed vs No response, P = 0, Suppressed vs Activated, P = 0, No response vs Activated, P = 0, Bonferroni corrected for multiple comparisons) indicating that our thresholds on neuropil subtracted traces reliably separate different classes of trial-wise responses even in the absence of neuropil-subtraction. **d,** Example raw (light red, light blue) and neuropil-subtracted (dark red, dark blue) traces for neurons reliably activated (red) and suppressed (blue) across trials (using neuropil-subtracted trial-wise magnitudes). Neuropil subtraction slightly decreases response magnitudes in both groups, though doesn't induce large changes in response time-course. Negative responses can be seen with and without neuropil-subtraction. Shading is mean ± s.e.m across neurons. Traces have been linearly interpolated across the photostimulation epoch (grey box)

to avoid the photostimulation artefact for display only. All box-plots are median (mid-line) with 25th and 75th percentiles (box) and 5th and 95th percentiles (whiskers). $N = 3.68 \times 10^6$ trial-wise responses, 30894 neurons, 11 sessions, 6 mice, 1 – 2 sessions each.

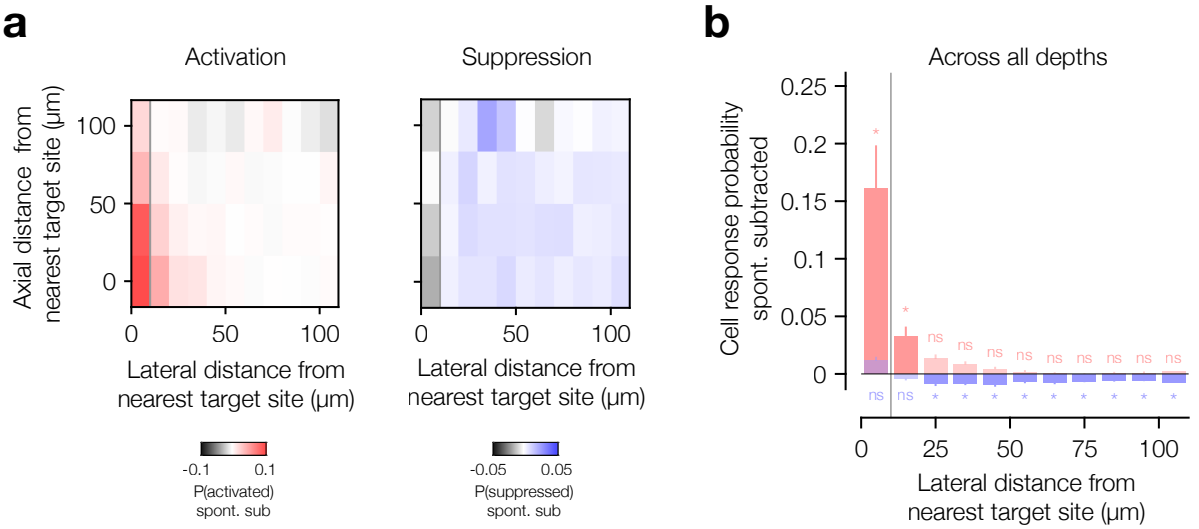

Figure 3 – figure supplement 3 | Activation and suppression have different spatial profiles. **a,** Lateral and axial spatial profile of activation (*left*) and suppression (*right*) relative to nearest target site co-ordinate: for each neuron plot its distance to nearest target site co-ordinate vs the proportion of trials it was activated or suppressed. Bin spatial distances and then average across all neurons in each bin. Average this value across stimulus trial types. **b,** Quantification of spread of activation and suppression with lateral distance, collapsed across all axial planes. All data in all panels are hit:miss matched (see **Figure 3 – figure supplement 1** and *Methods*) and have had catch trial values subtracted. All tests in *b* are Wilcoxon signed-rank tests, Bonferroni corrected for the number of lateral bins. * denotes $P < 0.05$. $N = 10$ sessions, 6 mice, 1 – 2 sessions each. Note one session was excluded as there were no hits on catch trials so we were unable to subtract a hit:miss matched neural response rate on catch trials from the stimulus trial data. Error bars are s.e.m. and spatial bin widths are 10 µm.

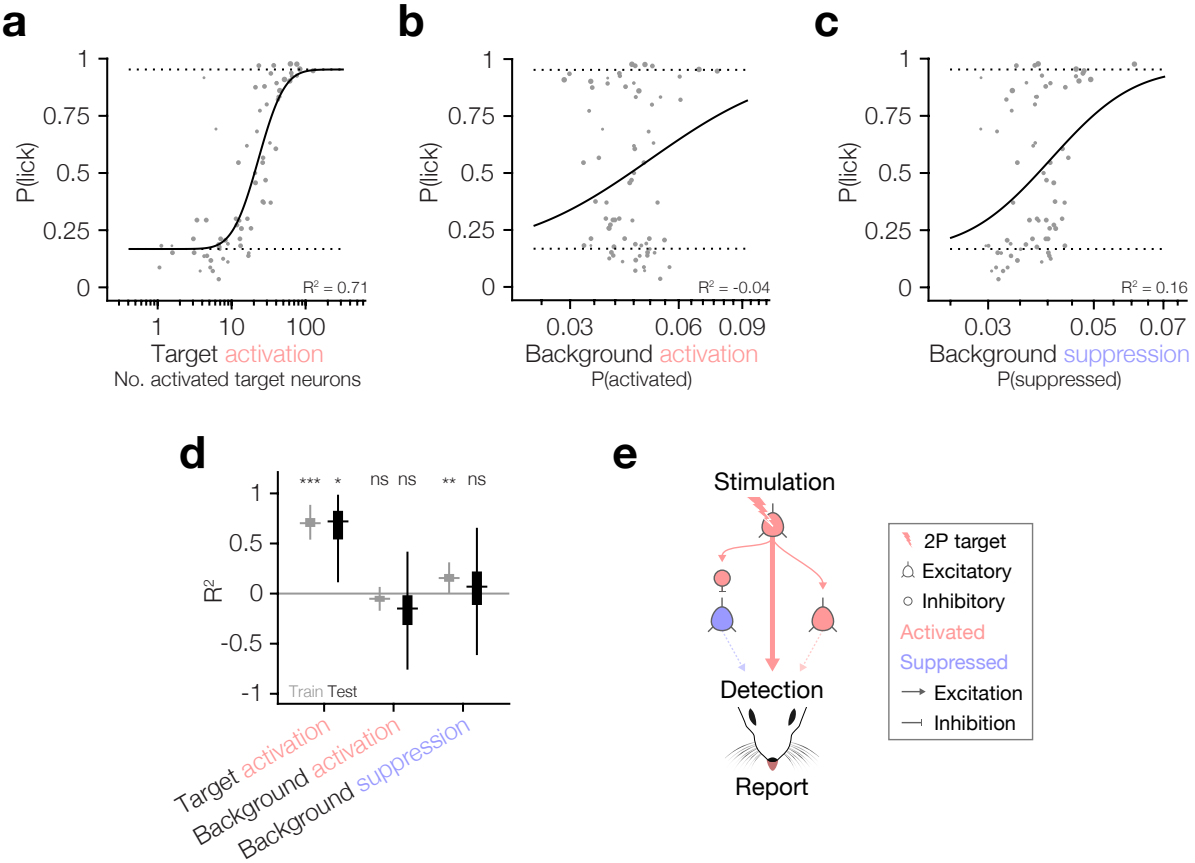

Figure 4 | Behaviour follows the activity of targeted ensembles despite matched suppression in the local network.

a – c, Psychometric curve fits relating behavioural detection to the number of targets activated (*a*), the proportion of the background network activated (*b*) and the proportion of the background network suppressed (*c*). Solid lines: psychometric curve fit; dotted lines: fixed lapse rate (upper) and false alarm rate (bottom) for psychometric fit. Note that all neural data has been hit:miss matched (see *Methods*) so the effective P(lick) for all datapoints is 0.5, however for each datapoint we fit the actual recorded P(lick). The similarity of panel *a*, which is hit:miss matched, to **Figure 2d**, which isn't, demonstrates that the contribution of lick signals to the relationship between target activation and behaviour is negligible. Fits and R^2 values reported are quantified on all data (compared to cross-validated fits in following panels). For these panels $N = 11$ sessions, 6 animals, 1 – 2 sessions each. Some trial types from some sessions are excluded for having too few hits or misses to be able to match the hit:miss ratio. **d,** Variance explained (R^2) by the three predictors in *a* – *c* during the training (grey) and testing (black) phases of cross-validation (10,000 permutations of 80:20 train:test split). Only target activation strongly and reliably explains behaviour across both training and testing. Background suppression is mildly predictive of behaviour in model training datasets, but this relationship does not generalise to test datasets. Background activation does not explain behaviour. Boxplots are median with 25th/75th percentile boxes and whiskers extending to the most extreme data points not considered outliers (see *Methods*). **e,** Schematic summarising the three tested routes from cortical activation to behavioural report highlighting that only the activity of target neurons has any reliable influence on behaviour despite matched suppression in the local network.