

Inhibitory engrams in perception and memory.

Helen Barron^{1,2*}, Tim P. Vogels^{3*}, Timothy Behrens^{1,4*}, and Mani Ramaswami^{5,6*}

¹The Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, Oxford OX3 9DU, UK. ²Medical Research Council Brain Network Dynamics Unit, Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK. ³Centre for Neural Circuits and Behaviour, University of Oxford, Oxford OX1 3SR, UK. ⁴The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3BG, UK. ⁵Trinity College Institute of Neuroscience, School of Genetics and Microbiology and School of Natural Sciences/ Trinity College Dublin, Dublin-2 Ireland. ⁶National Centre for Biological Sciences, TIFR Centre, Bangalore 560065, India.

Running Title: Excitation Inhibition Balancing in Memory and Cognition

***Correspondence:** helen.barron@merton.ox.ac.uk; behrens@fmrib.ox.ac.uk; tim.vogels@cncb.ox.ac.uk; mani.ramaswami@tcd.ie

Keywords: Neural Representation, Negative Image, Computational, Model, Drosophila, Human, Attention, Neuromodulation, Neural network.

SUMMARY

Nervous systems use excitatory cell assemblies or "perceptual engrams" to encode and represent sensory percepts. Similarly, synaptically connected cell assemblies or "memory engrams" are thought to represent memories of past experience. Multiple lines of recent evidence indicate that brain systems also create inhibitory replicas of excitatory engrams with important cognitive functions. Such matched inhibitory engrams may form through homeostatic potentiation of inhibition onto postsynaptic cells that show increased levels of excitation. Inhibitory engrams can reduce behavioral responses to familiar stimuli thereby resulting in behavioral habituation. In addition, by preventing inappropriate activation of excitatory memory engrams, inhibitory engrams can make memories quiescent, stored in a latent form that is available for context-relevant activation. In neural networks with balanced excitatory and inhibitory engrams, the release of innate responses and recall of associative memories can occur through focussed disinhibition. Understanding mechanisms that regulate the formation and expression of inhibitory engrams *in vivo* may help not only to explain key features of cognition, but also to provide insight into transdiagnostic traits associated with psychiatric conditions such as autism, schizophrenia and post-traumatic stress disorder (PTSD).

\body

Percepts and memories are thought to be represented in the brain by excitatory activity in groups of neurons, described as cell-assemblies (Hebb, 1949) or memory engrams (Semon, 1921). Excitatory activity within a cell-assembly thus allows percepts to be represented, while coordinated activity across different cell-assemblies facilitates formation and recall of associative memories. These excitatory engrams are stored in quiescent form that allows them to be reactivated in appropriate contexts. Several recent studies indicate that the storage and reactivation of excitatory engrams may be respectively accompanied by creation and modulation of matched inhibitory engrams. Such inhibitory engrams, also termed "negative images" or "inhibitory representations," can be constructed in neural networks through simple, evolutionarily primitive, synaptic and cellular mechanisms, which are recognized to be involved in the phenomenon of excitatory-inhibitory balance observed across the brain.

Neurons and neural circuits normally operate within a preset range of activity. Outside this range, altered levels of neuronal spiking trigger a variety of homeostatic mechanisms ranging from compensatory ion-channel expression to local synaptic scaling (Nelson and Turrigiano, 2008; Turrigiano, 2012). These homeostatic mechanisms ensure that the activity parameters of a neuron operate around a set-point such that the spiking rates remain stable and a balance of depolarizing and hyperpolarizing currents is maintained. This ensures that a homeostatic equilibrium is preserved, where excitation and inhibition are both locally and globally balanced, despite plastic changes across neurons and synapses .

A potentially important candidate among these regulatory homeostatic mechanisms is potentiation of inhibitory synapses, which occurs under specific conditions and acts to prevent excessive levels of postsynaptic activity (D'Amour J and Froemke, 2015; Das et al., 2011; Fischer and Carew, 1993; Froemke et al., 2007; Vogels et al., 2011). An underappreciated property of this balancing process is that it naturally creates "inhibitory representations" or "negative images" of new, unbalanced excitatory patterns that arise within neural networks in response to experience (Barron et al., 2016a; Ramaswami, 2014). Here, we discuss how such inhibitory engrams created

through compensatory inhibitory potentiation contribute to cognitive function, and ask how their disturbance could contribute to clinical features and transdiagnostic traits observed in neuropsychiatric conditions. In doing so, we integrate insights from diverse neurophysiological, behavioral, computational and brain imaging studies across multiple species, to extract and highlight fundamental principles of memory consolidation and recall (Barron et al., 2016a; Das et al., 2011; Kato et al., 2015; Letzkus et al., 2015; Ramaswami, 2014; Vogels et al., 2011).

Direct evidence for restoration of EI balance through inhibitory synapse potentiation.

Excitation and inhibition appear balanced in cortical neurons, both at a global (Moore and Nelson, 1998; van Vreeswijk and Sompolinsky, 1996; Wehr and Zador, 2003) and local level (Froemke et al., 2007; Okun and Lampl, 2008; Vogels and Abbott, 2009). At steady state, cortical neurons therefore show closely matched depolarising and hyperpolarising currents (Okun and Lampl, 2008; Wehr and Zador, 2003). However, this balance is disturbed during learning when excitatory plasticity occurs following transient reduction in inhibition (Vogels and Abbott, 2009) (Chen et al., 2015; Kruglikov and Rudy, 2008; Letzkus et al., 2015). For subsequent storage of learned information, the balance between excitation and inhibition must therefore be restored. One means by which this is achieved is via compensatory inhibitory synaptic potentiation (Froemke et al., 2007; Kuchibhotla et al., 2016) (Figure 1).

At an electrophysiological level, this process of compensatory inhibitory synaptic potentiation has been best characterized in the mammalian auditory cortex. Principal cells in the auditory cortex normally show balanced depolarising and hyperpolarising currents across a range of tonal frequencies but each cell is optimally tuned to a preferred frequency at which maximal excitatory and inhibitory postsynaptic currents (EPSCs and IPSCs) occur (Wehr and Zador, 2003). Following a simple form of learning where exposure to a specific tone is paired with direct stimulation of nucleus basalis (NB), a structure that mediates attentional engagement, a shift in the preferred frequency of a principal neuron towards the frequency of the exposed tone can be observed (Froemke et al., 2007) (Figure 1b). Initial changes during this “representational plasticity” occur during NB stimulation due to the reduction in feed-forward inhibition following acetylcholine release (Chen et al., 2015; Sur et al.,

2013). Therefore, a shift in the preferred tonal frequency is first observed in the excitatory post-synaptic currents (EPSCs), disrupting the balance between excitation and inhibition at the new frequency (Froemke et al., 2007). Strikingly, if tone exposure is continued for 120-180 minutes without associated NB-stimulation, then a rebalancing process occurs. This rebalancing can be attributed to potentiation of the IPSC, which shifts to match the EPSC with a peak at the new frequency (Figure 1c). The mechanism responsible for this EI-balance restoration has since been investigated by both computational and experimental approaches.

Computational simulations indicate that simple synaptic plasticity rules are sufficient to account for rebalancing. In a model network of postsynaptic, integrate and fire neurons, Vogels et al. applied a simple spike-time dependent plasticity rule that acts on inhibitory synapses to mediate inhibitory-to-excitatory connections (Vogels et al., 2011). The plasticity rule allows the strength of inhibitory synapses to increase if inhibitory neurons fire within 40 msec of their excitatory postsynaptic cell, and vice versa. When this rule is implemented within an unbalanced systems where excitatory synapses are stronger than their inhibitory counterparts, simulations show that inhibitory synaptic potentiation occurs until EPSCs and IPSCs are precisely matched. With some tuning of the target spiking rate for the postsynaptic neuron, the learning rate and the spiking frequency of inhibitory neurons, the experimentally observed phenomenon of EI rebalancing can be accurately reproduced (Vogels et al., 2011) (Figure 1c).

Experimental observations and theoretical models therefore agree that inhibitory potentiation plays a critical role in rebalancing cortical networks following excitatory plasticity. At an empirical level, inhibitory potentiation is relatively poorly characterised and underlying mechanisms could vary across brain regions and inhibitory cell types. Moreover, alternative strategies, such as changes in excitatory drive (Banerjee et al., 2016), may also contribute to cortical rebalancing. Similarly, at a theoretical level, the precise learning rule responsible for inhibitory potentiation remains open to debate (Vogels et al., Frontiers 2013) as the implementation of alternative synaptic learning rules can also successfully account for cortical rebalancing (Luz and Shamir, 2012). Nevertheless, empirical investigations in rodent auditory cortex emphasise the importance of inhibitory potentiation as a means to

rebalance neural circuits, and show that near coincident pre- and post-synaptic activity is critical to the learning rule (D'Amour J and Froemke, 2015). Furthermore, inhibitory potentiation is reported to be dependent upon NMDA receptors, suggesting that NMDA receptors act as coincidence detectors to trigger retrograde signals to active presynaptic neurons, thereby coordinating plasticity between co-active excitatory and inhibitory neurons (Nugent and Kauer, 2008; Nugent et al., 2007; Sudhakaran et al., 2012; Woodin et al., 2003). Thus, by simulating biologically plausible learning rules for inhibitory potentiation, theoretical models can explain key features of EI balancing observed during representational plasticity in the auditory cortex (Figure 1c).

Negative images in adaptive stimulus filtering and behavioral habituation

Following increased activity across an ensemble of excitatory synapses, the form of EI balancing described above is predicted to result in delayed strengthening of matched inhibitory synapse ensembles (Barron et al., 2016a; Ramaswami, 2014) (Figure 2A-C). This creates inhibitory engrams or representations, which counterbalance excitatory representations and ensure that EI balance is maintained in face of increased excitation. In large-scale neural networks, such inhibitory engrams may underlie multiple fundamental cognitive processes. We first consider their role in habituation, a form of non-associative, implicit memory that reduces innate responses to irrelevant stimuli (Ramaswami, 2014; Rankin et al., 2009; Thompson and Spencer, 1966).

To explain olfactory habituation in *Drosophila*, behavioral genetic analyses have inferred and invoked an inhibitory learning rule similar to that involved in restoring EI balance in mammalian auditory cortex, (Das et al., 2011; Larkin et al., 2010; Ramaswami, 2014). In insects, odorants are encoded by assemblies of projection neurons in the antennal lobe, a structure homologous to the mammalian olfactory bulb. Many lines of evidence argue that in a neutral environment, prolonged activation of an odorant-specific excitatory assembly results in the selective strengthening of inhibitory synapses onto neurons activated by the odorant (Das et al., 2011; Glanzman, 2011). This results in the formation of an inhibitory replica of the specific pattern of odor-induced excitation. The newly created inhibitory

representation of odorant-induced excitation acts as a filter to specifically attenuate physiological and behavioural responses to the familiar and inconsequential odorant (Ramaswami, 2014). Significantly, as observed for the EI balancing process associated with re-tuning cells in mammalian auditory cortex, insect olfactory habituation also requires postsynaptic NMDA receptors suggesting that a common homeostatic mechanism is at play (D'Amour J and Froemke, 2015; Das et al., 2011). Together, these observations suggest that olfactory habituation may be usefully considered to arise through a form of EI balancing in which inhibitory potentiation serves to restrain the spiking of a subset of odorant-activated projection neurons to within a behaviorally appropriate range.

Broadly similar mechanisms for auditory habituation have been recently reported in the mammalian cortex, albeit without direct evidence for the underlying synaptic mechanism (Kato et al., 2015) (Fig 2D-L). Here, *in vivo* GCaMP-based imaging shows that tone-specific auditory habituation is associated with reduced calcium-fluxes in layer 2/3 pyramidal cells in the rodent auditory cortex. This reduction in pyramidal cell activity is accompanied by enhanced activity of somatostatin positive (SOM) inhibitory neurons in the same brain region. Habituation can thus be characterised as a 10-fold reduction in the excitation/inhibition ratio of population activity. Interestingly, when attention to the tone becomes important for task performance, SOM inhibition is reduced and pyramidal cell responses to the tone increase even in habituated animals (Kato et al., 2015).

Taken together, data from insect and mammalian nervous systems suggest that habituation may generally arise through the formation of inhibitory engrams created by inhibitory synaptic potentiation via mechanisms that resemble those involved in EI balancing (D'Amour J and Froemke, 2015; Froemke et al., 2007; Kato et al., 2015; Ramaswami, 2014; Vogels et al., 2011). We note that in the mammalian brain, the subtypes of inhibitory interneurons involved and their mode of regulation may depend on the particular neural circuit in which they are embedded. For instance, while SOM-positive interneurons have been identified as playing a critical role in auditory habituation, PV-positive interneurons are noted for their more pervasive role in equalising the ratio between inhibition and excitation (Xue et al., 2014). Although the precise interaction between these different subtypes is not yet clear, the model in

which matched inhibition drives habituation generates important predictions. One notable corollary suggests that innate behavioural responses attenuated by habituation can later be rapidly restored through disinhibition, via inputs that silence the relevant inhibitory neurons (Kato et al., 2015). This provides an explanation for the psychological phenomena of dishabituation and attentional override, two defining properties of habituation (Ramaswami, 2014; Rankin et al., 2009; Thompson and Spencer, 1966).

Regulating the formation of inhibitory engrams

As evidenced from studies of representational plasticity and habituation discussed above, both the formation and activity of inhibitory engrams are regulated by behavioral context. But under what circumstances do these inhibitory engrams form? By definition, behavioural habituation occurs to non-salient and inconsequential stimuli. Indeed, contextual inputs that confer salience or assign emotional significance to a stimulus are known to actively block habituation (Ramaswami, 2014; Rankin et al., 2009; Thompson and Spencer, 1966). This is probably best revealed by physiological studies investigating regulation of intrinsic inhibitory synaptic plasticity (metaplasticity) in the circuit that underlies habituation of a siphon withdrawal reflex in *Aplysia* (Bristol and Carew, 2005; Fischer et al., 1997). When multiple mild stimuli are applied to the tail, a reduced tail-touch induced siphon withdrawal is observed as a consequence of potentiated inhibitory feedback onto siphon motoneurons. By contrast, a single electric shock applied to the tail increases tail-touch induced sensitized siphon withdrawal through serotonin-dependent potentiation of excitatory sensorimotor synapses. Remarkably, the tail-shock induced release of serotonin not only potentiates excitatory connections but also blocks habituation by reducing the ability of repeated, mild tactile tail stimulation to cause inhibitory interneuron-motoneuron synapse facilitation (Bristol and Carew, 2005; Fischer et al., 1997) (Figure 3).

Interestingly, in the mammalian auditory cortex, inhibitory potentiation following EPSC enhancement occurs in response to continued tonal stimulation, but this has only been observed in the absence of cholinergic NB stimulation (Froemke et al., 2007). As cholinergic NB afferents mediate disinhibition (Froemke, 2007), this leads

to the prediction that concurrent NB stimulation would actively inhibit heterosynaptic inhibitory plasticity to prevent EI rebalancing. More generally, we speculate that when memories are actively encoded during transient periods of high neuromodulator concentration, EI rebalancing mechanisms may be disrupted. Abnormal persistence of such neuromodulation may result in so-called maladaptive memories that persistently reactivate, as observed in post-traumatic stress disorder (Ehlers and Clark, 2000). In conclusion, processes that underlie the restoration of EI balance in both the retuning of cells in the mammalian auditory cortex and in behavioural habituation best described in invertebrates, not only both rely on inhibitory plasticity, but may also be regulated in a similar manner by context-dependent neuromodulation.

Inhibitory engrams in associative memory storage and recall

In addition to the proposed role for inhibitory potentiation in cellular re-tuning and habituation, recent experiments in humans suggest that inhibitory potentiation may also play a fundamental role in the storage and recall of associative memories (Barron et al., 2016a) (Figure 4). Using representational fMRI (Barron et al., 2016b), the expression of associative memories can be indexed immediately after learning, within the region of cortex that typically represents the stimuli in question. For example, when participants learn to associate rotationally-invariant abstract shapes, representational overlap is observed in an anterior region of the lateral occipital cortex (LOC). However, over time, this cortical expression of the associative memories decreases, such that it becomes invisible to fMRI 24 hours after learning. Remarkably, when the level of the cortical inhibitory transmitter GABA is reduced using anodal transcranial direct current stimulation (tDCS), the expression of the associative memory reappears. Thus, associative memory transitions from an early form which is visible to fMRI, to a later quiescent state that can only be revealed under conditions of reduced inhibition. This result is consistent with the idea that excitatory synaptic potentiation that occurs during learning, is later matched by equivalent inhibitory synaptic potentiation and the formation of a matching inhibitory representation.

Together these observations point to the following model (Figure 4). Excitatory potentiation during learning enhances connectivity between distinct neuronal ensembles that encode the associated stimuli in question. EI balance is then restored

by inhibitory synaptic potentiation, which at a network level creates an inhibitory engram of the newly strengthened excitatory connections to match and cancel/ reduce their effect. Once the balanced state is restored, weakly connected excitatory neurons are balanced by weak inhibition, while strongly connected excitatory neurons are balanced by strong inhibition. Thus, strong associations are stored in a latent state that allows them to be selectively revealed by disinhibition (Barron et al., 2016a).

These findings are consistent with electrophysiological recording in a songbird while it listens to a tutor's song (Figure 4). When juvenile zebra finches are exposed to a tutor song, excitatory neurons in HVC are active. However, in adult birds HVC-neuron responses to tutor song are silenced by inhibitory inputs, whose strength is proportional to the relative similarity of the birds own song to the tutors (Vallentin et al., 2016). Disinhibiting the HVC via application of a GABA agonist, Gabazine, releases excitation to reveal a response not dissimilar from that observed in developing juvenile birds. This suggests that inhibition plays an important role in protecting stored representations from interference with closely related information.

The temporal period and the precise mechanisms involved in the inferred EI balancing process are not known. But a particularly attractive notion is that this occurs during sleep, potentially during sharp-wave ripples, when previous experiences and memories are replayed offline (Eschenko et al., 2008; Ji and Wilson, 2007; Ramadan et al., 2009). We speculate that this could be one of the ways in which sleep contributes not only to memory consolidation, but also to homeostatic processes that prepare the brain for new learning.

Within this framework, it remains unproven as to how stored information is normally recalled. While mechanisms for memory recall could vary across brain systems and circuits, we suggest cortical disinhibition as one potential strategy to allow release of strongly connected excitatory ensembles from balanced strong inhibition (Kato et al., 2015; Vallentin, 2016; Barron et al., 2016a). Disinhibition may be mediated by neuromodulators, acting for instance through muscarinic acetylcholine receptors on cortical interneurons (Kim et al., 2016). More selective disinhibition may be additionally mediated through neural structures critical for memory recall, such as the hippocampus. Thus, in addition to the established function of local disinhibition to

enhance excitatory transmission (Vogels and Abbott, 2009) and contribute to the initial encoding of memory (Chen et al., 2015; Kruglikov and Rudy, 2008; Letzkus et al., 2015), disinhibition may also play a second significant function to facilitate release of previously learned, but latent cortical associations (Barron et al., 2016a).

Experimental evidence for disinhibition in memory recall is particularly evident following fear conditioning, the most simple form of associative memory (Courtin et al., 2014; Letzkus et al., 2015; Wolff et al., 2014). Here, when a conditioned auditory stimulus (CS) is used to trigger fear-memory recall in rats, phasic inhibition of a subset of parvalbumin-positive (PV) GABAergic interneurons can be observed in dorso-medial prefrontal cortex (Courtin et al., 2014). This higher level inhibition of inhibitory neurons can be conceived to bring about disinhibition of the learned fear response. Indeed, optogenetic inhibition of these GABAergic neurons is sufficient to induce freezing (Courtin et al., 2014). This disinhibition may be mediated by vasointestinal peptide positive (VIP) interneurons (David et al., 2007; Pfeffer et al., 2013; Pi et al., 2013), although details of the disinhibitory elements and their circuitry can vary between different brain regions and types of associative learning (Fu et al., 2014; Wolff et al., 2014).

A framework for normal and variant memory storage and recall

The observations and arguments above lead to a common systems level framework for memory storage and recall that are built on two strikingly simple principles (Figure 5). (1) Following habituation or associative training respectively, innate behaviors and simple associative memories are masked by inhibitory engrams. These compensatory inhibitory representations ensure that EI balance is maintained despite new learning and may be considered a critical component of memory consolidation. During habituation, inhibitory engrams are created when stimuli are experienced without concurrent engagement of emotional or attentional circuits (Figures 1-2, Figure 5).__Following associative learning, we speculate that sharp-wave ripple dependent replay during sleep may play an important role in their formation. (2) Innate behaviours and associative memories can be recalled in appropriate contexts through selective and local disinhibition. We speculate that local disinhibition is driven by contextual or attentional inputs that recruit secondary inhibitory neurons to selectively target cell-types encoding inhibitory engrams (Chen et al., 2015; Letzkus

et al., 2015; Perisse et al., 2016) (Figure 2; Figure 4; Figure 5). In this manner volitional and directed recall of memories may occur via sensory stimulation or attentional mechanisms, which act to disinhibit and activate cortical subdomains that store specific memories. Although we focus here on inhibitory engrams created by local EI balancing mechanisms, the framework we propose also naturally accommodates inhibitory representations constructed through potential alternative mechanisms (Clark, 2013; Cooke et al., 2015; Kogo and Trengove, 2015; Ramaswami, 2014; Rao and Ballard, 1999; Wacongne et al., 2012).

The model appears to have considerable explanatory and heuristic value. For instance, it conceptually distinguishes between behavioral amnesia arising from defects in storage from those arising from defects in recall. It is therefore relevant to recent work which shows that a memory engram remains preserved in amnesic rats that do not express contextual fear memory (Roy et al., 2016; Ryan et al., 2015). Consistent with the framework we present here, in the mouse models of Alzheimer's disease, memory storage largely remains intact while recall is disrupted, potentially by a failure of disinhibitory pathways that mediate recall of stored memories (Figure 5).

Implications for EI disruption in clinical conditions

Many investigators have worked towards establishing a common framework to account for the behavioural phenotypes common to psychiatric conditions, which otherwise have complex and distinct genetic bases (Lisman et al., 2008; Markram et al., 2007; Oberman et al., 2005; Sahin and Sur, 2015; Walsh et al., 2008; Yizhar et al., 2011). Our premise that a major function of EI balancing is to mask irrelevant perceptions, memories and behavior, allows mechanistic predictions to be drawn that concern specific cognitive dysfunctions that may arise when the process is disrupted (Figure 5). These predictions are particularly relevant, given several lines of genetic and physiological evidence pointing to an imbalance between excitation and inhibition as a possible substrate for symptoms observed in a range of clinical conditions including autism and schizophrenia (Chou et al., 2010; Gogolla et al., 2009; Han et al., 2014; Lisman et al., 2008; Nelson and Valakh, 2015; Penagarikano et al., 2011; Rubenstein and Merzenich, 2003; Tuchman et al., 2010; Tyzio et al., 2014; Yizhar et al., 2011).

One predicted consequence of defects in the formation of inhibitory engrams is weak behavioral habituation that could arise either from increased neuromodulatory activity which blocks the formation of inhibitory representations, or from defects in inhibitory function and plasticity. Remarkably, weak habituation is particularly common feature of autism spectrum disorders (ASD) that is also observed in schizophrenia (Ethridge et al., 2016; Guiraud et al., 2011; Kleinhans et al., 2009; Rankin et al., 2009; Sinha et al., 2014). Weak habituation could also result in several downstream cognitive changes observed in autism and in schizophrenia, including altered sensory gating, stimulus hypersensitivity and reduced ability to cope in complex environments, where multiple signals may appear salient and compete for attention (Ethridge et al., 2016; Ramaswami, 2014; Rankin et al., 2009; Sinha et al., 2014; Vogels and Abbott, 2007). It is also conceivable that additional features of autism such as sticky attention, wherein familiar stimuli remain engaging for unusually long periods of time while novel stimuli seem challenging, could arise from weak habituation and hypersensitivity to novel stimuli (Landry and Bryson, 2004; Sinha et al., 2014; Vogels and Abbott, 2007). An additional predicted consequence of defects in inhibitory rebalancing that we postulate to be necessary for masking associative memories, is strong and contextually inappropriate recall of associative memories. Strikingly, unusually strong and vivid associative memories are often associated with ASD (Zamoscik et al., 2016). It is conceivable that vivid recall of unconnected memory traces could lead to inappropriate associations and thought disturbances and delusions observed in schizophrenia (Figure 5). This is consistent with theoretical models in which EI imbalance in neural networks has been used to account for hallucinations and delusional experiences that may be observed both in autism and in schizophrenia (Gao and Penzes, 2015; Toal et al., 2009; Vogels and Abbott, 2007).

This mechanistic proposal that specific cognitive features and transdiagnostic traits of autism and schizophrenia arise in part due to defects in the creation of inhibitory engrams is now experimentally testable, taking advantage of MRI paradigms that provide an index for EI balance across associative memories (Barron et al., 2016b; Barron et al., 2016a). One simple testable prediction is that EI imbalance across cortical associations will persist for unusually long time periods in affected individuals. While the validity of this proposal will require additional investigation, we suggest

that a model that connects synaptic mechanisms of EI balancing in neural circuits with fundamental cognitive processes altered in psychiatric conditions may provide a unifying principle to explain common disturbances that arise from diverse molecular and genetic perturbations.

Conclusion

In summary, the model that inhibitory engrams created by EI balancing mechanisms can reversibly suppress and thereby allow context-relevant recall of innate and learned behaviors provides an elegant mechanistic framework to understand fundamental cognitive functions in both health and disease. Specifically, this framework provides a simple description for key processes that underlie memory storage and recall in psychotypical cognitive function, which gives insight into the substrates of neuropsychiatric symptoms that emerge when EI balance is disturbed.

ACKNOWLEDGEMENTS

We thank Jens Hillebrand for help with the figures, Isabell Twick for comments on the manuscript, Kevin Mitchell, Srikanth Ramaswamy and Mriganka Sur for useful discussions. H.C.B. is supported by a Junior Research Fellowship from Merton College, University of Oxford, and the John Fell Oxford University Press (OUP) Research Fund (153/046). T.P.V. is funded by a Sir Henry Dale Fellowship from the Wellcome Trust (WT100000). T.E.J.B. is supported by a Wellcome Trust Senior Research Fellowship (WT104765MA) together with a James S. McDonnell Foundation Award (JSMF220020372). M.R. acknowledges grants from Science Foundation Ireland and the Simons Foundation Autism Research Initiative as well as core support from the NCBS, Tata Institute of Fundamental Research, Bangalore for collaborative work with K. VijayRaghavan.

REFERENCES

Banerjee, A., Rikhye, R.V., Breton-Provencher, V., Tang, X., Li, C., Li, K., Runyan, C.A., Fu, Z., Jaenisch, R., and Sur, M. (2016). Jointly reduced inhibition and excitation underlies circuit-wide changes in cortical processing in Rett syndrome. *Proc Natl Acad Sci U S A* *113*, E7287-E7296.

- Barron, H.C., Garvert, M.M., and Behrens, T.E. (2016b). Repetition suppression: a means to index neural representations using BOLD? *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 371.
- Barron, H.C., Vogels, T.P., Emir, U.E., Makin, T.R., O'Shea, J., Clare, S., Jbabdi, S., Dolan, R.J., and Behrens, T.E. (2016a). Unmasking Latent Inhibitory Connections in Human Cortex to Reveal Dormant Cortical Memories. *Neuron* 90, 191-203.
- Bristol, A.S., and Carew, T.J. (2005). Differential role of inhibition in habituation of two independent afferent pathways to a common motor output. *Learn Mem* 12, 52-60.
- Chen, N., Sugihara, H., and Sur, M. (2015). An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity. *Nature neuroscience* 18, 892-902.
- Chou, Y.H., Spletter, M.L., Yaksi, E., Leong, J.C., Wilson, R.I., and Luo, L. (2010). Diversity and wiring variability of olfactory local interneurons in the *Drosophila* antennal lobe. *Nature neuroscience*.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36, 181-204.
- Cooke, S.F., Komorowski, R.W., Kaplan, E.S., Gavornik, J.P., and Bear, M.F. (2015). Visual recognition memory, manifested as long-term habituation, requires synaptic plasticity in V1. *Nature neuroscience* 18, 262-271.
- Courtin, J., Karalis, N., Gonzalez-Campo, C., Wurtz, H., and Herry, C. (2014). Persistence of amygdala gamma oscillations during extinction learning predicts spontaneous fear recovery. *Neurobiology of learning and memory* 113, 82-89.
- D'Amour J, A., and Froemke, R.C. (2015). Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex. *Neuron* 86, 514-528.
- Das, S., Sadanandappa, M.K., Dervan, A., Larkin, A., Lee, J.A., Sudhakaran, I.P., Priya, R., Heidari, R., Holohan, E.E., Pimentel, A., *et al.* (2011). Plasticity of local GABAergic interneurons drives olfactory habituation. *Proc Natl Acad Sci U S A* 108, E646-654.
- David, C., Schleicher, A., Zuschratter, W., and Staiger, J.F. (2007). The innervation of parvalbumin-containing interneurons by VIP-immunopositive interneurons in the primary somatosensory cortex of the adult rat. *Eur J Neurosci* 25, 2329-2340.
- Ehlers, A., and Clark, D.M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour research and therapy* 38, 319-345.
- Eschenko, O., Ramadan, W., Molle, M., Born, J., and Sara, S.J. (2008). Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. *Learn Mem* 15, 222-228.
- Ethridge, L.E., White, S.P., Mosconi, M.W., Wang, J., Byerly, M.J., and Sweeney, J.A. (2016). Reduced habituation of auditory evoked potentials indicate cortical hyper-excitability in Fragile X Syndrome. *Translational psychiatry* 6, e787.
- Fischer, T.M., Blazis, D.E., Priver, N.A., and Carew, T.J. (1997). Metaplasticity at identified inhibitory synapses in *Aplysia*. *Nature* 389, 860-865.

- Fischer, T.M., and Carew, T.J. (1993). Activity-dependent potentiation of recurrent inhibition: a mechanism for dynamic gain control in the siphon withdrawal reflex of *Aplysia*. *J Neurosci* *13*, 1302-1314.
- Froemke, R.C., Merzenich, M.M., and Schreiner, C.E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature* *450*, 425-429.
- Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., and Stryker, M.P. (2014). A cortical circuit for gain control by behavioral state. *Cell* *156*, 1139-1152.
- Gao, R., and Penzes, P. (2015). Common mechanisms of excitatory and inhibitory imbalance in schizophrenia and autism spectrum disorders. *Current molecular medicine* *15*, 146-167.
- Glanzman, D.L. (2011). Olfactory habituation: fresh insights from flies. *Proc Natl Acad Sci U S A* *108*, 14711-14712.
- Gogolla, N., Leblanc, J.J., Quast, K.B., Sudhof, T.C., Fagiolini, M., and Hensch, T.K. (2009). Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *Journal of neurodevelopmental disorders* *1*, 172-181.
- Guiraud, J.A., Kushnerenko, E., Tomalski, P., Davies, K., Ribeiro, H., and Johnson, M.H. (2011). Differential habituation to repeated sounds in infants at high risk for autism. *Neuroreport* *22*, 845-849.
- Han, S., Tai, C., Jones, C.J., Scheuer, T., and Catterall, W.A. (2014). Enhancement of Inhibitory Neurotransmission by GABAA Receptors Having alpha2,3-Subunits Ameliorates Behavioral Deficits in a Mouse Model of Autism. *Neuron* *81*, 1282-1289.
- Hebb, D.O. (1949). *The Organization of Behavior: A Neuropsychological Theory* (New York: Wiley and Sons.).
- Ji, D., and Wilson, M.A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience* *10*, 100-107.
- Kato, H.K., Gillet, S.N., and Isaacson, J.S. (2015). Flexible Sensory Representations in Auditory Cortex Driven by Behavioral Relevance. *Neuron* *88*, 1027-1039.
- Kim, J.H., Jung, A.H., Jeong, D., Choi, I., Kim, K., Shin, S., Kim, S.J., and Lee, S.H. (2016). Selectivity of Neuromodulatory Projections from the Basal Forebrain and Locus Coeruleus to Primary Sensory Cortices. *J Neurosci* *36*, 5314-5327.
- Kleinmans, N.M., Johnson, L.C., Richards, T., Mahurin, R., Greenon, J., Dawson, G., and Aylward, E. (2009). Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. *Am J Psychiatry* *166*, 467-475.
- Kogo, N., and Trengove, C. (2015). Is predictive coding theory articulated enough to be testable? *Front Comput Neurosci* *9*, 111.
- Kruglikov, I., and Rudy, B. (2008). Perisomatic GABA release and thalamocortical integration onto neocortical excitatory cells are regulated by neuromodulators. *Neuron* *58*, 911-924.
- Kuchibhotla, K.V., Gill, J.V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A., Miller, K.D., and Froemke, R.C. (2016). Parallel processing by cortical inhibition enables context-dependent behavior. *Nature neuroscience*.

- Landry, R., and Bryson, S.E. (2004). Impaired disengagement of attention in young children with autism. *Journal of child psychology and psychiatry, and allied disciplines* *45*, 1115-1122.
- Larkin, A., Karak, S., Priya, R., Das, A., Ayyub, C., Ito, K., Rodrigues, V., and Ramaswami, M. (2010). Central synaptic mechanisms underlie short-term olfactory habituation in *Drosophila* larvae. *Learn Mem* *17*, 645-653.
- Letzkus, J.J., Wolff, S.B., and Luthi, A. (2015). Disinhibition, a Circuit Mechanism for Associative Learning and Memory. *Neuron* *88*, 264-276.
- Lisman, J.E., Coyle, J.T., Green, R.W., Javitt, D.C., Benes, F.M., Heckers, S., and Grace, A.A. (2008). Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. *Trends in neurosciences* *31*, 234-242.
- Luz, Y., and Shamir, M. (2012). Balancing feed-forward excitation and inhibition via Hebbian inhibitory synaptic plasticity. *PLoS computational biology* *8*, e1002334.
- Markram, H., Rinaldi, T., and Markram, K. (2007). The intense world syndrome--an alternative hypothesis for autism. *Front Neurosci* *1*, 77-96.
- Moore, C.I., and Nelson, S.B. (1998). Spatio-temporal subthreshold receptive fields in the vibrissa representation of rat primary somatosensory cortex. *J Neurophysiol* *80*, 2882-2892.
- Nelson, S.B., and Turrigiano, G.G. (2008). Strength through diversity. *Neuron* *60*, 477-482.
- Nelson, S.B., and Valakh, V. (2015). Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. *Neuron* *87*, 684-698.
- Nugent, F.S., and Kauer, J.A. (2008). LTP of GABAergic synapses in the ventral tegmental area and beyond. *The Journal of physiology* *586*, 1487-1493.
- Nugent, F.S., Penick, E.C., and Kauer, J.A. (2007). Opioids block long-term potentiation of inhibitory synapses. *Nature* *446*, 1086-1090.
- Oberman, L.M., Hubbard, E.M., McCleery, J.P., Altschuler, E.L., Ramachandran, V.S., and Pineda, J.A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res Cogn Brain Res* *24*, 190-198.
- Okun, M., and Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature neuroscience* *11*, 535-537.
- Penagarikano, O., Abrahams, B.S., Herman, E.I., Winden, K.D., Gdalyahu, A., Dong, H., Sonnenblick, L.I., Gruver, R., Almajano, J., Bragin, A., *et al.* (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* *147*, 235-246.
- Perisse, E., Oswald, D., Barnstedt, O., Talbot, C.B., Huetteroth, W., and Waddell, S. (2016). Aversive Learning and Appetitive Motivation Toggle Feed-Forward Inhibition in the *Drosophila* Mushroom Body. *Neuron* *90*, 1086-1099.
- Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., and Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience* *16*, 1068-1076.

- Pi, H.J., Hangya, B., Kvitsiani, D., Sanders, J.I., Huang, Z.J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature* 503, 521-524.
- Ramadan, W., Eschenko, O., and Sara, S.J. (2009). Hippocampal sharp wave/ripples during sleep for consolidation of associative memory. *PloS one* 4, e6697.
- Ramaswami, M. (2014). Network plasticity in adaptive filtering and behavioral habituation. *Neuron* 82, 1216-1229.
- Rankin, C.H., Abrams, T., Barry, R.J., Bhatnagar, S., Clayton, D.F., Colombo, J., Coppola, G., Geyer, M.A., Glanzman, D.L., Marsland, S., *et al.* (2009). Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. *Neurobiology of learning and memory* 92, 135-138.
- Rao, R.P., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* 2, 79-87.
- Roy, D.S., Arons, A., Mitchell, T.I., Pignatelli, M., Ryan, T.J., and Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature* 531, 508-512.
- Rubenstein, J.L., and Merzenich, M.M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav* 2, 255-267.
- Ryan, T.J., Roy, D.S., Pignatelli, M., Arons, A., and Tonegawa, S. (2015). Memory. Engram cells retain memory under retrograde amnesia. *Science* 348, 1007-1013.
- Sahin, M., and Sur, M. (2015). Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. *Science* 350.
- Semon, R. (1921). Chapter II. Engraphic Action of Stimuli on the Individual. . In *The Mneme*, T.b.L. Simon, ed. (London: George Allen and Unwin).
- Sinha, P., Kjelgaard, M.M., Gandhi, T.K., Tsourides, K., Cardinaux, A.L., Pantazis, D., Diamond, S.P., and Held, R.M. (2014). Autism as a disorder of prediction. *Proc Natl Acad Sci U S A* 111, 15220-15225.
- Sudhakaran, I.P., Holohan, E.E., Osman, S., Rodrigues, V., Vijayraghavan, K., and Ramaswami, M. (2012). Plasticity of recurrent inhibition in the *Drosophila* antennal lobe. *J Neurosci* 32, 7225-7231.
- Sur, M., Nagakura, I., Chen, N., and Sugihara, H. (2013). Mechanisms of plasticity in the developing and adult visual cortex. *Prog Brain Res* 207, 243-254.
- Thompson, R.F., and Spencer, W.A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review* 73, 16-43.
- Toal, F., Bloemen, O.J., Deeley, Q., Tunstall, N., Daly, E.M., Page, L., Brammer, M.J., Murphy, K.C., and Murphy, D.G. (2009). Psychosis and autism: magnetic resonance imaging study of brain anatomy. *The British journal of psychiatry : the journal of mental science* 194, 418-425.
- Tuchman, R., Cuccaro, M., and Alessandri, M. (2010). Autism and epilepsy: historical perspective. *Brain Dev* 32, 709-718.

- Turrigiano, G. (2012). Homeostatic synaptic plasticity: local and global mechanisms for stabilizing neuronal function. *Cold Spring Harbor perspectives in biology* 4, a005736.
- Tyzio, R., Nardou, R., Ferrari, D.C., Tsintsadze, T., Shahrokhi, A., Eftekhari, S., Khalilov, I., Tsintsadze, V., Brouchoud, C., Chazal, G., *et al.* (2014). Oxytocin-mediated GABA inhibition during delivery attenuates autism pathogenesis in rodent offspring. *Science* 343, 675-679.
- Vallentin, D., Kosche, G., Lipkind, D., and Long, M.A. (2016). Neural circuits. Inhibition protects acquired song segments during vocal learning in zebra finches. *Science* 351, 267-271.
- van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724-1726.
- Vogels, T.P., and Abbott, L.F. (2007). Gating deficits in model networks: a path to schizophrenia? *Pharmacopsychiatry* 40 Suppl 1, S73-77.
- Vogels, T.P., and Abbott, L.F. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nature neuroscience* 12, 483-491.
- Vogels, T.P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* 334, 1569-1573.
- Wacongne, C., Changeux, J.P., and Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci* 32, 3665-3678.
- Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., *et al.* (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320, 539-543.
- Wehr, M., and Zador, A.M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* 426, 442-446.
- Wolff, S.B., Grundemann, J., Tovote, P., Krabbe, S., Jacobson, G.A., Müller, C., Herry, C., Ehrlich, I., Friedrich, R.W., Letzkus, J.J., *et al.* (2014). Amygdala interneuron subtypes control fear learning through disinhibition. *Nature* 509, 453-458.
- Woodin, M.A., Ganguly, K., and Poo, M.M. (2003). Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl⁻ transporter activity. *Neuron* 39, 807-820.
- Xue, M., Atallah, B.V., and Scanziani, M. (2014). Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* 511, 596-600.
- Yizhar, O., Fenno, L.E., Prigge, M., Schneider, F., Davidson, T.J., O'Shea, D.J., Sohal, V.S., Goshen, I., Finkelstein, J., Paz, J.T., *et al.* (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* 477, 171-178.

Zamoscik, V., Mier, D., Schmidt, S.N., and Kirsch, P. (2016). Early Memories of Individuals on the Autism Spectrum Assessed Using Online Self-Reports. *Frontiers in psychiatry* 7, 79.

FIGURE LEGENDS.

Figure 1: An EI rebalancing process observed in rat auditory cortex can be computationally simulated by implementing inhibitory plasticity rules

A. A schematic canonical circuit diagram showing excitatory neurons in red and inhibitory neurons in blue. The postsynaptic excitatory neuron receives balanced excitatory and inhibitory inputs such that the inhibitory and excitatory currents may be considered to be tuned to the same frequency. **B-C. Top panel:** schematic circuit diagram showing the effect of plasticity, initially leading to EI imbalance following excitatory plasticity (**B**), and subsequent restoration of balance following inhibitory plasticity (**C**). **B-C. Middle panel:** *In-vivo* whole-cell recording in primary auditory cortex with EPSCs shown in red and IPSCs shown in blue. Concurrent cholinergic stimulation and exposure to a tone (black arrow) that is shifted relative to the original preferred frequency of the neuron (white arrow), modifies the EPSC such that it retunes and peaks at the frequency of the exposed tone (**B**). After repetitive tonal stimulation (in the absence of cholinergic activity) the IPSC eventually shifts from the original to the new frequency to match and rebalances the modified EPSC (**C**). **B-C. Lower panel:** Simulation from a neural network model where the excitatory tuning curve (red circles) is manually changed and a spike-time dependent inhibitory plasticity rule then applied. After 30 minutes of simulation the inhibitory tuning curve (blue squares) still shows the original tuning (**B**), however after 180 minutes it has shifted to match and rebalance the excitatory tuning (**C**). Data panels are adapted from Froemke et al., 2007 and Vogels et al, 2011.

Figure 2: Inhibitory representations can mediate sensory habituation

A-C. Schematic model for habituation based on (Ramaswami, 2014). **A.** A balanced circuit of inhibitory (blue) neurons and excitatory (red) neurons showing a connected excitatory representation. **B.** Continuous exposure to the relevant percept for the excitatory representation referred to in *A* results in EI imbalance, due to sustained excitatory activity (bright red). **C.** EI rebalancing through inhibitory potentiation creates a matched inhibitory representation (bright blue) to suppress the excitatory representation. **D-I.** Direct evidence for inhibitory potentiation during habituation from mouse auditory cortex (adapted from Kato et al. 2015). **D.** Habituation

protocol: mice were passively exposed to a tone for 5 days. Mice experienced 200 trials per day and on each trial the tone lasted for 5 to 9 seconds. **E-F.** In-vivo two-photon image of tone-evoked GCaMP6s-expressing neurons in layer 2/3 of auditory cortex on day 1 (**E**), and on day 5 (**F**). Excitation was sparser day 5. **G.** Canonical cortical microcircuit showing SOM-positive inhibitory neurons connecting to layer 2/3 (L2/3) excitatory neurons. **H-I.** The daily change index, used to assess the change in activity across days 1-5, showed a reduced in excitation in layers 2/3 (**H**) which was accompanied by an increase in excitation in SOM-positive inhibitory neurons (**I**). Together these results suggest a reduction in excitation and an increase in inhibition following habituation. **J-L.** In habituated animals, tone-relevant task engagement is accompanied by disinhibition of L2/3 excitatory cells (adapted from Kato et al. 2015). **J.** Tone-relevant task protocol: Habituated animals were trained to lick a food port in response to the habituated tone. The response was then compared to passive exposure to the habituated tone, in the absence of the food port. **K-L.** The change index reflects the increase in the layer 2/3 neuron excitation during the tone-relevant task, relative to the passive exposure. A task associated increase in the excitatory response was observed (**K**), which was accompanied by a decrease in tone-evoked activity in SOM-positive neurons (**L**), consistent with attentional disinhibition.

Figure 3: Neuromodulation gates inhibitory potentiation and EI balancing.
(Adapted from Fischer et al., 1997).

A. Simplified circuitry that mediates the tail-siphon withdrawal (T-SW) response in *Aplysia*. Sensory neurons (SN) in the tail activate an excitatory interneuron L29 (shown in pink) which excites both the siphon motor neuron (MN) and the inhibitory interneuron L30 (shown in blue). L30 mediates feedback inhibition onto L29. Tail shock results in serotonin release on both L30-L29 and sensorimotor synapses. T-SW habituation is accompanied by enhanced L30-L29 transmission (increased IPSCs recorded in L29), which can also be induced by direct L30 stimulation at frequencies normally induced by tail touch (Fischer, 1997; Bristol et al., 2005). **B-C.** Brief tetanic stimulation was delivered to the inhibitory interneuron L30 (shaded bar indicates stimulation period). The IPSC was recorded from the excitatory interneuron L29. Following stimulation, the IPSC from L29 was enhanced (control condition, open

circles). However, this enhancement was not observed if the tetanic stimulation to the inhibitory interneuron L30 was delivered 90 seconds after a tail-shock (**B**, filled squares, * indicates a significant difference of $p < 0.05$ relative to the control condition). Furthermore, this enhancement was also not observed if the tetanic stimulation to the inhibitory interneuron L30 was delivered 90 seconds after application of serotonin (**C**, filled squares, * indicates a significant difference of $p < 0.05$ relative to the control condition). Together these results suggest that inhibitory potentiation is similarly blocked by either tail-shock or direct application of serotonin. In other scenarios, neuromodulators could act by prevent inhibitory neuron firing.

Figure 4: Inhibitory rebalancing allows associative information to lie dormant unless EI balance is disturbed

A-C. Schematic representation of a neural network that stores two distinct cell assemblies which are represented by an ensemble of excitatory (either red or orange circles) and inhibitory (either blue or green squares). Initially the two cell assemblies are distinct, with one responding to the red input, and the other responding to the orange input (**A**). When the two cell assemblies become associated via excitatory plasticity between the two excitatory ensembles (black lines), co-activation is observed such that the red cell assembly responds to the orange input and the orange cell assembly responds to the red input (**B**). If inhibitory plasticity then acts to restore balance in the network, the newly strengthened excitatory connections are quenched by inhibition such that co-activation between the two cell assemblies is reduced (**C**).

D. In the human brain, representational fMRI can be used to index co-activation of neural representations for associated abstract shapes in an anterior region of the lateral occipital complex. However, this co-activation is only observed immediately after learning (timepoint 1), and decreases over time (timepoint 2 (~1 hour after learning) and timepoint 3 (~24 hours after learning)). Once the associative memory is quiescent, application of anodal transcranial direct current stimulation leads to a significant reduction in cortical GABA and an increase in the co-activation index for associated stimuli. This suggests that associative memories are stored in cortex in balanced excitatory-inhibitory ensembles, which lie dormant unless the balance between excitation and inhibition is disrupted. **E.** The increase in the co-activation index between associated representations can be predicted by the drop in GABA

induced by brain stimulation. **F.** In juvenile songbirds, premotor neurons show an increase in activity during exposure to a tutor song. **G.** In adult songbirds, premotor neurons are quiescent during exposure to a tutor song. **H.** Application of a GABA agonist, Gabazine, releases excitation in premotor neurons of the songbird, revealing a response not dissimilar from that observed in developing juvenile birds.

Figure 5: Inhibitory engrams in perception and memory: a model.

A hypothetical framework showing how inhibitory engrams could form in neural systems illustrated with reference to their influence in perceptual habituation, memory storage and recall. Excitatory inputs onto an array of postsynaptic neurons are shown as positive red peaks and inhibitory inputs as negative blue peaks. In the illustration above, these constitute excitatory and inhibitory engrams respectively.

Before habituation, sensory stimuli activate excitatory perceptual engrams but trigger relatively weak or imprecise inhibition. Repeated excitatory engram stimulation, with minimal attentional or emotional engagement, results in formation of a matched inhibitory engram that reduces the stimulus response and causes behavioral habituation (Ramaswami, 2014). Attentional inputs or dishabituating stimuli can restore the initial stimulus response by promoting disinhibitory inputs that suppress inhibitory engrams (Ramaswami, 2014; Kato et al., 2015).

Similarly, memory is first encoded as excitatory engrams. Over time, matched inhibitory engrams form to rebalance the neural network (Barron et al., 2016; Valentin et al., 2016). The formation of these inhibitory memory engrams may occur via homeostatic mechanisms similar to those involved in habituation, potentially during replay of excitatory memory engrams. Once formed, inhibitory engrams allow memories to be stored in a dormant form for context-appropriate recall, which we hypothesise to occur through focussed disinhibition.

(Below in green) Defects in inhibitory engram formation or regulation are predicted to cause perceptual and cognitive abnormalities including transdiagnostic traits observed in autism, schizophrenia and post-traumatic stress disorders.

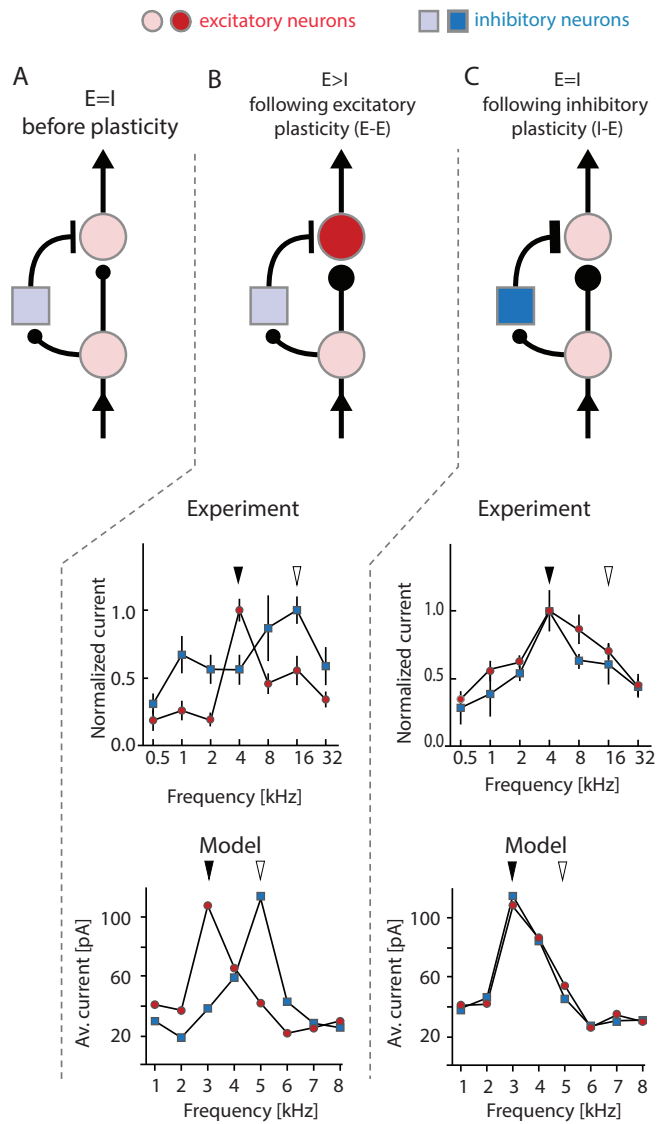


Figure 1

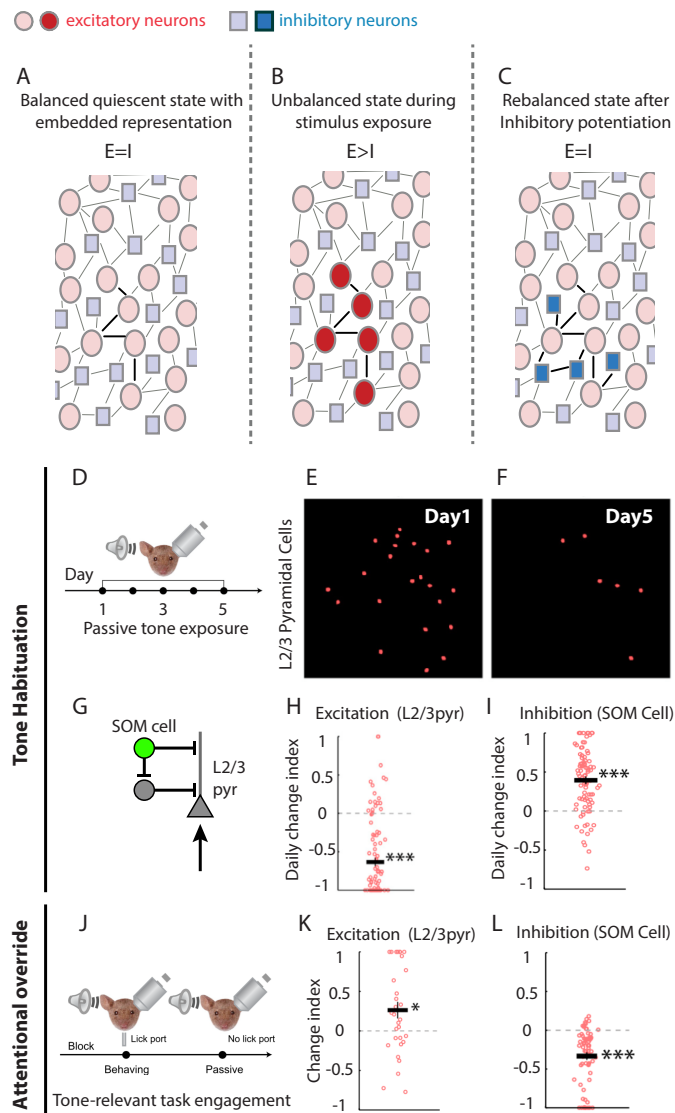


Figure 2

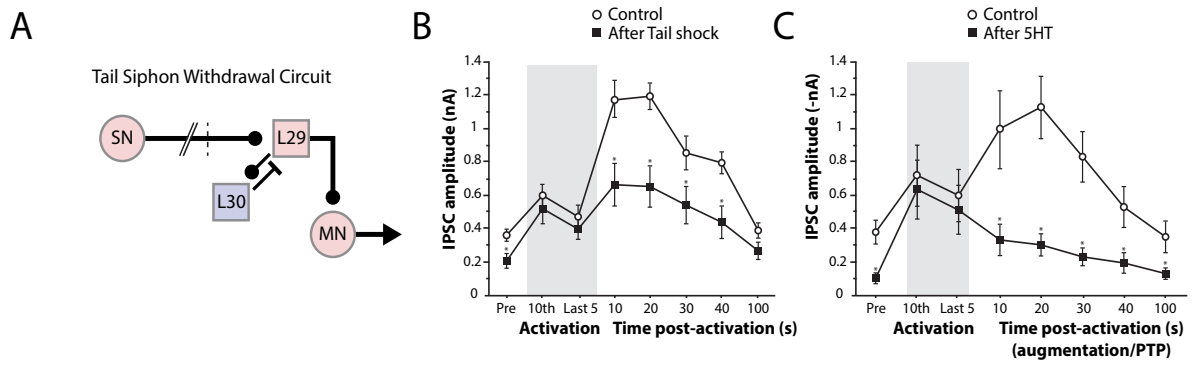


Figure 3

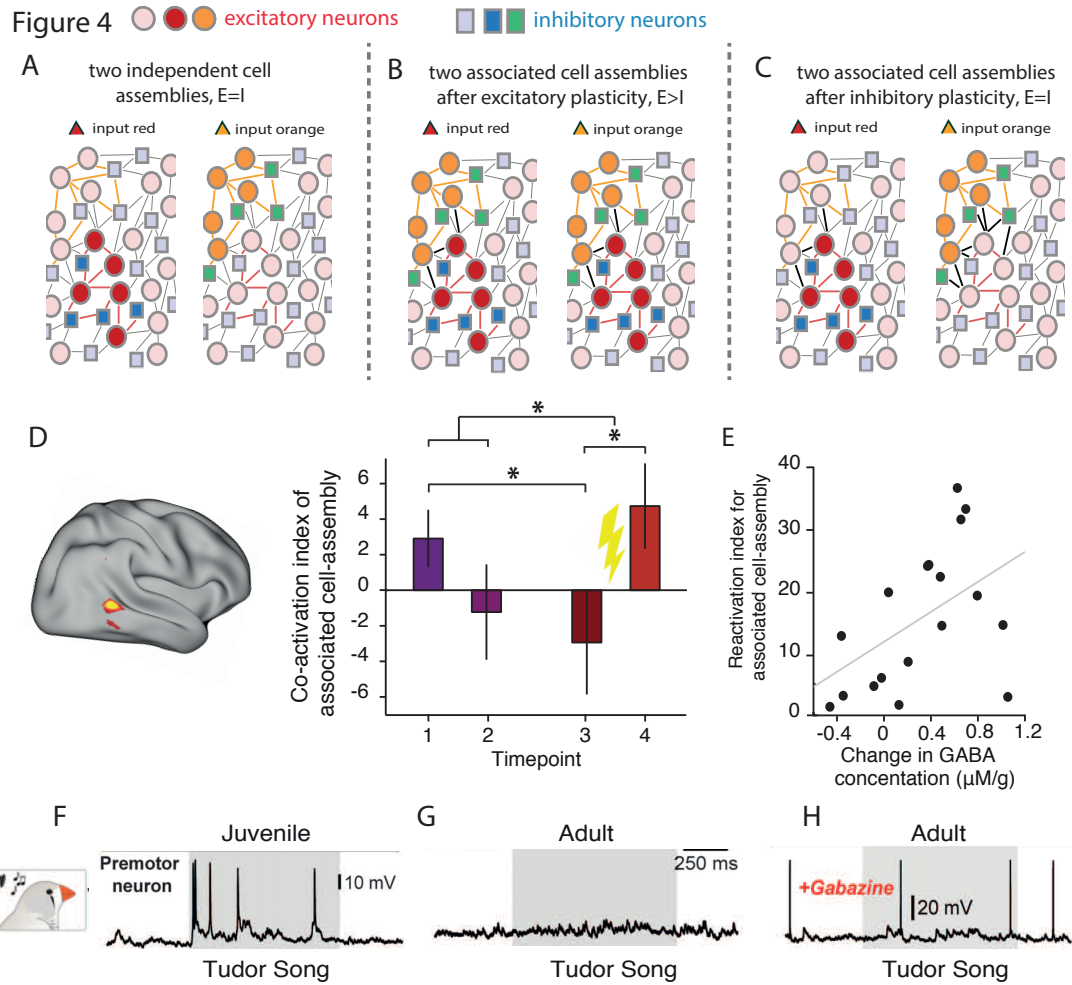


Figure 4

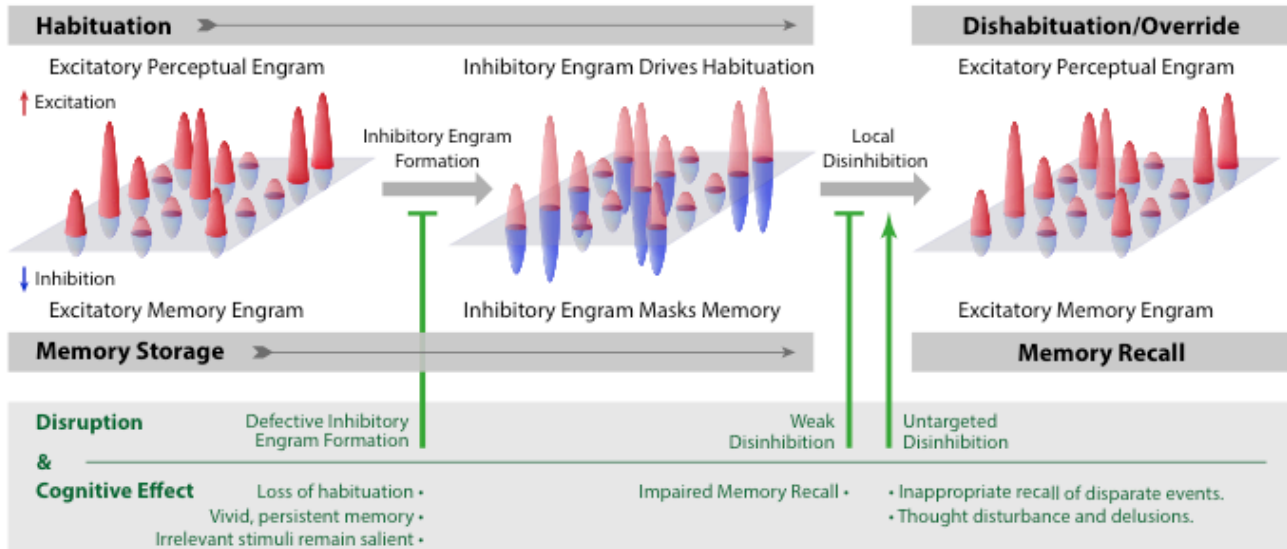


Figure 5