Cerebral Cortex Advance Access published November 29, 2012

Cerebral Cortex
doi:10.1093/cercor/bhs370

Linking Microcircuit Dysfunction to Cognitive Impairment: Effects of Disinhibition Associated with Schizophrenia in a Cortical Working Memory Model

John D. Murray[1,2], Alan Anticevic[3,4,5], Mark Gancsos[3], Megan Ichinose[3], Philip R. Corlett[3,5], John H. Krystal[3,4,5,6,7] and Xiao-Jing Wang[2,8]

[1]Department of Physics, Yale University, New Haven, CT 06520, USA, [2]Department of Neurobiology, Kavli Institute of Neuroscience, [3]Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06510, USA, [4]National Institute on Alcohol Abuse and Alcoholism (NIAAA), Center for the Translational Neuroscience of Alcoholism (CTNA), [5]Department of Psychiatry, Abraham Ribicoff Research Facilities, Connecticut Mental Health Center, Yale University, New Haven, CT 06519, USA, [6]Psychiatry Services, Department of Psychiatry, Yale-New Haven Hospital, New Haven, CT, USA, [7]Division of Clinical Neuroscience, Veterans Affairs (VA) National Center for Post-Traumatic Stress Disorder (PTSD), VA Connecticut Healthcare System, West Haven, CT 06516, USA and [8]Center for Neural Science, New York University, New York, NY 10003, USA

Address correspondence to Xiao-Jing Wang, Center for Neural Science, New York University, 4 Washington Place, New York, NY 10003, USA. Email: xjwang@nyu.edu

Excitation–inhibition balance (E/I balance) is a fundamental property of cortical microcircuitry. Disruption of E/I balance in prefrontal cortex is hypothesized to underlie cognitive deficits observed in neuropsychiatric illnesses such as schizophrenia. To elucidate the link between these phenomena, we incorporated synaptic disinhibition, via *N*-methyl-D-aspartate receptor perturbation on interneurons, into a network model of spatial working memory (WM). At the neural level, disinhibition broadens the tuning of WM-related, stimulus-selective persistent activity patterns. The model predicts that this change at the neural level leads to 2 primary behavioral deficits: 1) increased behavioral variability that degrades the precision of stored information and 2) decreased ability to filter out distractors during WM maintenance. We specifically tested the main model prediction, broadened WM representation under disinhibition, using behavioral data from human subjects performing a spatial WM task combined with ketamine infusion, a pharmacological model of schizophrenia hypothesized to induce disinhibition. Ketamine increased errors in a pattern predicted by the model. Finally, as proof-of-principle, we demonstrate that WM deteriorations in the model can be ameliorated by compensations that restore E/I balance. Our findings identify specific ways by which cortical disinhibition affects WM, suggesting new experimental designs for probing the brain mechanisms of WM deficits in schizophrenia.

Keywords: disinhibition, NMDAR hypofunction, prefrontal cortex, schizophrenia, working memory

Introduction

A basic principle of cortical computation is the dynamically balanced interaction of excitatory pyramidal cells and inhibitory interneurons (Shadlen and Newsome 1994; van Vreeswijk and Sompolinsky 1996; Shu et al. 2003). Disruption of the ratio of excitation and inhibition (E/I balance) can give rise to profound behavioral deficits and may play a key role in serious mental illnesses such as schizophrenia (Yizhar et al. 2011). Specifically, there is growing support for a hypothesis that cortical disinhibition occurs in schizophrenia, due to the disrupted functioning of inhibitory interneurons, which results in elevated E/I ratio (Lewis et al. 2005; Marin 2012). However, it remains unclear how disinhibition may produce deficits in higher cognition, which are a prominent feature of neuropsychiatric diseases.

Cognitive deficits are postulated to be at the "core" of schizophrenia neuropathology (Elvevåg and Goldberg 2000; Barch and Ceaser 2012), with a severe deficit in working memory (WM), the ability to transiently maintain and manipulate information internally (Goldman-Rakic 1994; Lee and Park 2005). In addition to maintenance over time, robust WM requires shielding internal representations from interference by both internal noise and external distraction. This "filtering" function, in particular, may be severely compromised in schizophrenia (Anticevic et al. 2011).

The prefrontal cortex is a key node in the distributed cortical network recruited during WM (Owen et al. 2005; Fuster 2008) and exhibits altered inhibitory microcircuitry in schizophrenia (Lewis et al. 2005; Bitanihirwe et al. 2009). Primate electrophysiological studies show that persistent firing of prefrontal pyramidal cells supports robust and stable WM representations (Funahashi et al. 1989) that depend critically on E/I balance (Rao et al. 2000). Biophysically realistic computational models have delineated a neural circuit basis of WM activity through 2 key network properties: Strong recurrent excitation to sustain persistent activity and recruitment of lateral inhibition to shape selectivity of representations (Compte et al. 2000; Brunel and Wang 2001).

An open question is whether cortical disinhibition, specifically within a WM microcircuit, can give rise to the types of deficits that may be observed in schizophrenia. To test this hypothesis, we examined the neural and behavioral effects of disinhibition in a spiking neural network model of spatial WM (Compte et al. 2000). Specifically, we induce disinhibition in the model via perturbation of *N*-methyl-D-aspartate receptors (NMDARs) on inhibitory interneurons, which weakens the recruitment of feedback inhibition. This mechanism may be linked to the pathophysiology of schizophrenia (Lisman et al. 2008; Nakazawa et al. 2012) and possibly accounts for some of the key effects of NMDAR antagonists, such as ketamine, which are a leading pharmacological model of schizophrenia (Krystal et al. 1994).

The model predicts that, at the neural level, disinhibition elevates baseline firing rates and broadens WM activity patterns. At the behavioral level, disinhibition increases deterioration over time of the precision of stored information during WM and increases the range of distractors that can interfere with WM. To experimentally investigate synaptic mechanisms and test the main model prediction of broadened WM

representations under disinhibition, we used behavioral data from healthy human subjects performing a spatial WM match/nonmatch task following ketamine administration. It was found that performance errors increased with ketamine specifically for those nontarget test probes that are similar to the remembered targets, as predicted by the model. Finally, as proof-of-principle, we demonstrate that synaptic compensations that restore E/I balance in the model can ameliorate the effects of disinhibition.

Materials and Methods

We used an extensively validated spiking network model of spatial WM, consisting of pyramidal cells and interneurons, structured in a ring architecture (Fig. 1*A*; Compte et al. 2000; Carter and Wang 2007). Parameters were modified starting from the original "modulated parameter set" of Compte et al. (2000) to produce 1) a narrower persistent activity pattern (adjusting the connectivity profile), 2) drift more comparable with experimental observations during human behavior (Badcock et al. 2008; adjusting background input), and 3) robustness of multistability to small (~1%) decreases in recurrent excitatory conductances (adjusting recurrent conductances). Notably, all reported effects were present in the original parameter set of Compte et al. (2000) and were quite robust to parameter modulations.

Single-Neuron Models

Pyramidal cells and interneurons are modeled as leaky integrate-and-fire neurons (Tuckwell 1988), characterized by total capacitance C_m, leak conductance g_L, resting potential V_L, spike threshold potential V_{th}, reset potential V_{res}, and refractory time τ_{ref}. For pyramidal cells, $C_m = 0.5$ nF, $g_L = 25$ nS, $V_L = -70$ mV, $V_{th} = -50$ mV, $V_{res} = -60$ mV, and $\tau_{ref} = 2$ ms; for interneurons, $C_m = 0.2$ nF, $g_L = 20$ nS, $V_L = -70$ mV, $V_{th} = -50$ mV, $V_{res} = -60$ mV, and $\tau_{ref} = 1$ ms. The subthreshold membrane potential, $V(t)$, follows:

$$C_m \frac{dV(t)}{dt} = -g_L(V(t) - V_L) - I(t), \quad (1)$$

where $I(t)$ is the total input current to the cell.

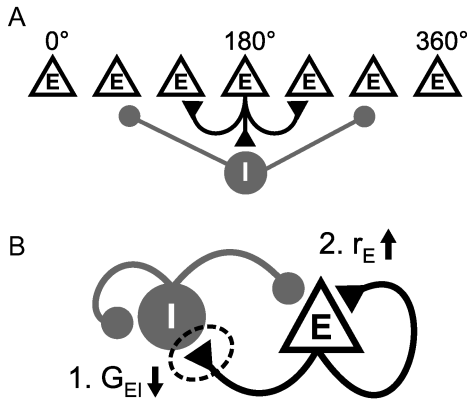

Figure 1. Recurrent model of spatial WM and disinhibition mechanism. (*A*) Schematic network architecture. The model consists of recurrently connected excitatory pyramidal cells (E) and inhibitory interneurons (I). Pyramidal cells are labeled by the angular location they encode (0–360°). Excitatory-to-excitatory connections are structured, such that neurons with similar preferred angles are more strongly connected. Connections between pyramidal cells and interneurons are unstructured and mediate feedback inhibition. (*B*) Disinhibition mechanism. NMDAR hypofunction on interneurons (1; decreased NMDAR conductance on interneurons, G_{EI}) weakens the recruitment of feedback inhibition. As a result, pyramidal cells are disinhibited and exhibit increased firing rates (2; increased firing rate of pyramidal cells, r_E).

Synaptic Interactions

The network consists of $N_E = 2048$ pyramidal cells and $N_I = 512$ inhibitory interneurons. Neurons receive recurrent, background, and external inputs. Excitatory synaptic currents are mediated by 2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid [AMPA] receptors (AMPARs) and NMDARs, and inhibitory synaptic currents are mediated by γ-aminobutyric acid type A ($GABA_A$) receptors (GABARs). The total input current to each neuron is:

$$I = I_{NMDA} + I_{AMPA} + I_{GABA} + I_{ext}, \quad (2)$$

where I_{ext} delivers stimulus input to pyramidal cells. The dynamics of synaptic currents for neuron i follow:

$$I_{i,AMPA} = (V_i - V_E) \sum_j g_{ji,AMPA} s_{j,AMPA}, \quad (3)$$

$$I_{i,NMDA} = (V_i - V_E) \frac{\sum_j g_{ji,NMDA} s_{j,NMDA}}{1 + [Mg^{2+}] \exp(-0.062\, V_i / 3.57\ \text{mV})}, \quad (4)$$

$$I_{i,GABA} = (V_i - V_I) \sum_j g_{ji,GABA} s_{j,GABA}, \quad (5)$$

where $V_E = 0$ mV and $V_I = -70$ mV, and $g_{ji,syn}$ denotes the synaptic conductance strength on neuron i from neuron j. NMDAR-mediated currents exhibit voltage dependence controlled by the extracellular magnesium concentration $[Mg^{2+}] = 1$ mM (Jahr and Stevens 1990).

Given a spike train $\{t_k\}$ in the presynaptic neuron j, the gating variables $s_{j,AMPA}$ and $s_{j,GABA}$ for AMPAR- and GABAR-mediated currents, respectively, are modeled as:

$$\frac{ds}{dt} = -\frac{s}{\tau_s} + \sum_k \delta(t - t_k). \quad (6)$$

The gating variable $s_{j,NMDA}$ for NMDAR-mediated current is modeled as:

$$\frac{dx}{dt} = -\frac{x}{\tau_x} + \sum_k \delta(t - t_k), \quad (7)$$

$$\frac{ds}{dt} = -\frac{s}{\tau_s} + \alpha_s x(1 - s), \quad (8)$$

with $\alpha_s = 0.5$ kHz and $\tau_x = 2$ ms. The decay time constant τ_s is 2 ms for AMPA, 10 ms for GABA, and 100 ms for NMDA. For simplicity, background inputs are mediated entirely by AMPARs, and recurrent excitatory inputs are mediated entirely by NMDARs, as they are critical for the stability of persistent activity (Wang 1999; Compte et al. 2000). Background input to each neuron is provided by an independent Poisson spike train with the rate of 0.6 kHz and AMPAR conductances of 9.3 nS on pyramidal cells and 7.14 nS on interneurons.

Network Model

As noted, pyramidal cells are organized in a ring architecture and are tuned to the angular location on a circle (0–360°), with uniform distribution of their preferred angles. The network structure follows a columnar architecture, such that pyramidal cells with similar stimulus selectivity are preferentially connected to each other. The synaptic conductance on neuron i from neuron j, $g_{ji,syn} = W(\theta_j - \theta_i) G_{syn}$, where θ_i is the preferred angle of neuron i, and $W(\theta_j - \theta_i)$ is the connectivity profile normalized such that:

$$\frac{1}{360^\circ} \int_{0^\circ}^{360^\circ} W(\theta) d\theta = 1. \quad (9)$$

For pyramidal-to-pyramidal connection, $W(\theta_j - \theta_i) = J^- + J^+ \exp[-(\theta_j - \theta_i)^2 / 2\sigma^2]$. We take $J^+ = 3$ and $\sigma = 9°$. All other synaptic connection profiles are unstructured. Synaptic conductance strengths are given by $G_{EE} = 1001.9/N_E$ nS, $G_{EI} = 717.6/N_E$ nS, $G_{IE} = 807.2/N_I$ nS, and $G_{II} = 566.2/N_I$ nS. The recurrent excitatory conductances, G_{EE} and G_{EI},

are parametrically reduced to study the effects of altered E/I ratio on neural activity and behavior.

Stimulus

As previously validated (Compte et al. 2000), we followed the stimulus protocol of Funahashi et al. (1989). Inputs are modeled as injected current with a Gaussian profile, $I(\theta) = I_0 \exp[-(\theta - \theta_c)^2/2\sigma_I^2]$, where the maximum current $I_0 = 375$ pA, θ_c is the stimulus location, and the width parameter $\sigma_I = 6°$. WM robustness to external disruption was tested with distractors. That is, we examined how stable the WM trace is, for control versus disinhibition conditions, when explicitly perturbed by an additional distracting input. Distractors were modeled identically to the initial cues, with same intensity and duration, but with a different stimulus location, such that the distractor appeared at a given angle relative to the original cue. Such a distractor manipulation allowed for the testing of specific hypotheses under disinhibition (discussed below).

Synaptic Manipulations

Disinhibition is implemented through a reduction of NMDAR conductance on interneurons, G_{EI} (Fig. 1*B*). For all simulations except those in Fig. 3, we make the approximation of reducing recurrent excitatory conductance only on interneurons, since that site is hypothesized to be the primary site of the action of NMDAR antagonists such as (Krystal et al. 2003; Kotermanski et al. 2009). To test the ability of compensations to restore E/I balance, we modeled 2 other synaptic manipulations: 1) the reduction of presynaptic glutamate release, by decreasing α_s in the glutamatergic synaptic update (Equation 8); 2) the enhancement of GABAR conductance on pyramidal cells, by increasing G_{IE}. If disinhibition results in elevated E/I ratio, the glutamatergic or GABAergic compensation can restore E/I balance by reducing excitation or increasing inhibition, respectively.

Simulation

Simulations were implemented with the Brian neural simulator (Goodman and Brette 2008). Simulation code is available from the authors upon request.

Analysis of Simulation Data

To decode the behavioral report from neural activity, we use the population vector approach (Georgopoulos et al. 1986): The angular location encoded by the network is given by $\theta_{report} = \arg\left[\sum_k r_k \exp(i\,\theta_k)\right]$. To visualize the scatter of WM reports in Figure 4*D*, we used the method described in Renart et al. (2003) to convert 1-dimensional (1D) angular position in the model into a 2-dimensional position. Population vectors were computed with a 50-ms temporal window. As done previously, spatiotemporal plots of network activity were smoothed over time (50 ms) and the pyramidal-cell population (40 neurons; Compte et al. 2000).

To measure the tuning width of the WM population activity pattern (bump attractor), we fit the firing rate profile of the population using a von Mises profile with sigmoidal saturation: $r(\theta) = r_0 + r_1[1 + \exp(-\beta(A\exp(\kappa\cos(2\pi(\theta - \theta_c)))))]^{-1}$. The bump width calculated as the full width at half maximum of the fit curve. Firing rate profile fits were computed with a 500-ms temporal window.

WM Task with Ketamine Manipulation

Complete experimental details are reported in Anticevic et al. (2012). Briefly, healthy human subjects performed a previously validated delayed spatial WM task (Leung et al. 2002; Driesen et al. 2008). Task details are described in Fig. 6*A*. Two or 4 targets are sequentially presented, followed by a 16-s delay period during which the target locations are to be held in WM. At the end of the delay, a probe is presented, either at a target or nontarget location, and subjects provided a button-press response of match or nonmatch, respectively. Subjects performed a series of such WM trials first during a session of saline infusion to act as a control, followed by a low-dose ketamine administration (achieved plasma concentration of 183 ng/mL (0.77 μM)) using the pharmacokinetic parameters of a 3-compartment model (Domino et al. 1982). Subjects completed a total 32 WM trials and 16 control task trials with no WM encoding and maintenance requirement but requiring a probe response, to control for motor effects per infusion (64 WM and 32 control trials in total). The experiment also included 2 additional randomized sessions in which subjects were pretreated with 2 dosages of a positive allosteric modulator of Group II metabotropic glutamate receptors (Patil et al. 2007). Behavior from these pretreatment sessions was not included in present analyses, as it does not directly relate to model predictions (due to multiple compound interactions), and will be reported elsewhere. As is standard statistical practice (Howell 2012), here, we report a simple effect for the factor involving no pretreatment, explicitly relevant to the model predictions.

Match/Nonmatch Simulation

To allow for a comparison between simulated and experimental ketamine data, we implemented a simple model of the comparison and decision computations involved in a WM match/nonmatch task (Engel and Wang 2011). For the firing rate profile of the WM bump, we use the fit function described above for calculating bump width. Next, we center that profile at the target location, removing the effect of drift. The probe representation is given by a Gaussian profile centered at the probe angle θ_p: $r_p(\theta) = \exp[-(\theta - \theta_p)^2/2\sigma_p^2]$, with $\sigma_p = 6°$. Following recent theoretical models (Tagamets and Horwitz, 1998; Engel and Wang 2011), the match/nonmatch decision is based on the overlap between target and probe representations. We calculate such overlap by taking the dot product of the 2 profiles. This overlap is normalized to be 1 when the probe is at the target location with control network parameters, and 0 when the probe is 180° away with control network parameters. To model the downstream decision circuit, the probability of match response, P_M, is then given by a sigmoidal function of the normalized overlap, x: $P_M(x) = p_0 + (p_1 - p_0)[1 + \exp(-(x - x_c)/\sigma)]^{-1}$. To replicate the experimental error rates, we set the parameters of the decision circuit to $p_0 = 0.18$, $p_1 = 0.78$, $x_c = 0.44$, and $\sigma = 0.075$.

Results

The network model, consisting of interconnected excitatory pyramidal cells and inhibitory interneurons, represents a local microcircuit in prefrontal cortex. The pyramidal cells exhibit selectivity to angular location on a circle (Fig. 1*A*; Funahashi et al. 1989). To subserve WM, the circuit exhibits multistability (Amit and Brunel 1997; Wang 1999; Compte et al. 2000). In addition to supporting a uniform, low-firing baseline state, the circuit allows for a continuous family of WM states that are supported by the strong recurrent excitation and capable of encoding stimulus location across a delay. A brief localized stimulus can activate a bell-shaped persistent activity pattern (bump attractor), consistent with location-selective sustained activity observed in prefrontal neurons during the delay period of WM tasks (Funahashi et al. 1989). Excitation in the network is dynamically balanced with inhibition, since increased pyramidal cell activity enhances the activity of interneurons that deliver feedback inhibition.

Disinhibition of Neural Activity

We implemented disinhibition in the model by reducing the strength of NMDAR conductance on interneurons, G_{EI} (Fig. 1*B*). Here, we set our disinhibition condition as G_{EI} reduced by 3.25% (for a complete parameter space characterization of our disinhibition manipulation, Fig. 3). Figure 2*A* shows spatiotemporal plots of the network activity during a

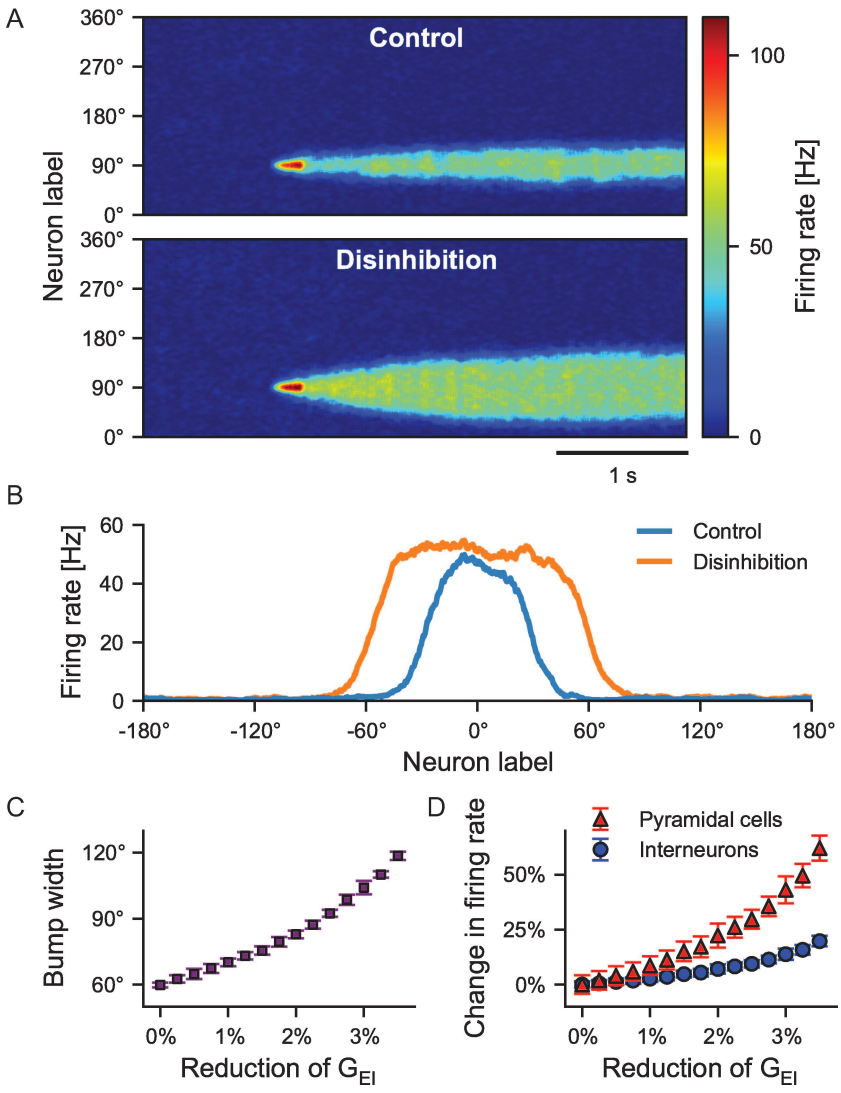

Figure 2. Disinhibition broadens WM representation. (*A*) The spatiotemporal plot of persistent activity patterns for control (upper panel) and disinhibition (3.25% reduction of G_{EI}) condition (lower panel). A stimulus is presented at 90° for 250 ms and, during the subsequent delay, the stimulus location is encoded by persistent activity of a WM bump. Disinhibition broadens the network activity pattern. (*B*) The firing rate profile of the WM bump for control and disinhibition conditions averaged over 500 ms in a single trial. (*C*) Increase of bump width as a function of the degree of disinhibition. (*D*) Increase of baseline firing rates as a function of the degree of disinhibition for both pyramidal cells and interneurons, from their control values of 1 and 6.4 Hz, respectively. Error bars show standard error of the mean.

WM trial for control and disinhibition conditions. Before the stimulus is applied, the network is in the uniform baseline state. A brief stimulus excites a subset of pyramidal cells selective to the stimulus angle. After removal of the stimulus, the stimulus angle is encoded throughout the delay by a persistent WM bump. Disinhibition substantially broadens the firing-rate profile of the bump (Fig. 2*B*,*C*). Because inhibitory interneurons are less strongly recruited by pyramidal-cell activity, more pyramidal cells can be activated by recurrent collaterals. Notably, network multistability is preserved at this level of disinhibition (i.e. the baseline state and WM bump states remain stable).

We also examined the effects of disinhibition on the spontaneous firing rate (Fig. 2*D*). As expected, disinhibition increases firing rates for pyramidal cells, since feedback inhibition is weakened. For strongly recurrent networks, excitation is dynamically balanced by inhibition via the recruitment of interneurons by pyramidal cells. Disinhibition also increases the firing rates of interneurons, though to a lesser degree than in pyramidal cells. In a strongly recurrent network such as ours, even though the strength of excitatory synaptic conductance to interneurons is reduced, their firing rates can still increase. This is because the disinhibited pyramidal cells increase their firing rates by a large enough proportion that consequently the net excitatory input to interneurons is also increased (Tsodyks et al. 1997).

Persistent WM activity patterns depend on both targets of recurrent excitatory synapses, interneurons, and pyramidal cells. Pyramidal-to-pyramidal synapses, with strength G_{EE}, are critical for the recurrent excitation necessary for persistent activity (Wang 1999). NMDAR antagonists likely act on both interneurons and pyramidal cells, although they may affect postsynaptic cell types differentially (Kotermanski and Johnson 2009). For the sake of parsimony in the model, all recurrent excitatory synapses are mediated by NMDARs. This approximation allows the direct comparison of the network's

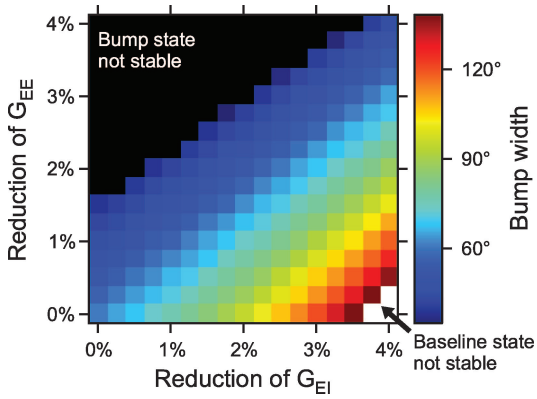

Figure 3. The parameter space of NMDAR hypofunction highlights the importance of E/I balance for multistability and bump width. Bump width increases with the reduction of G_{EI} and decreases with the reduction of G_{EE}. Bump width can be maintained with a proportional decrease in both G_{EI} and G_{EE}. The colored diagonal band is the region with multistability, in which the network supports both a low-rate symmetric baseline state and a family of WM bump states. In the white region at the bottom right corner, the baseline state is destabilized, due to strong disinhibition. In the black region at the top left corner, the bump state is destabilized, due to insufficient recurrent excitation.

sensitivity to changes in the net E/I ratio, but not the dissection of the relative roles of recurrent AMPARs and NMDARs. The tolerance of the model to NMDAR perturbations would be greater if a large fraction of the recurrent excitatory conductance was mediated by AMPARs rather than NMDARs. This is important in light of recent studies showing that the AMPAR/NMDAR ratio differs between pyramidal cells and fast-spiking interneurons (Wang and Gao 2009; Rotaru et al. 2011). To explicitly examine the model's sensitivity to E/I ratio, we parametrically decreased both G_{EI} and G_{EE}, while testing multistability and measuring bump width (Fig. 3).

Here, we found that if G_{EI} and G_{EE} are reduced together in a particular proportion to each other, the bump width is unaltered, that is, E/I balance is maintained. This suggests that E/I ratio is the critical parameter for network function, rather than the absolute strength of either G_{EI} or G_{EE} alone. If G_{EI} is reduced in greater proportion, then the network is in a state of disinhibition, and the WM bump width is increased (lower right corner in Fig. 3). In contrast, if G_{EE} is reduced in greater proportion, the bump width is decreased (upper left corner in Fig. 3). These bump-width effects are present within the parameter region of network multistability around a diagonal line in the parameter space of G_{EI} and G_{EE}. Taken together, these results show that the net effect of mild reductions in excitatory drive on neural WM activity can be described by the net change to E/I balance.

If E/I imbalance is substantial, either too elevated or too reduced, we observe a loss of network multistability. If disinhibition is too strong (via elevated E/I ratio), then the spontaneous state is no longer stable, and multistability is lost (lower right corner of Fig. 3). In that regime, a WM bump spontaneously appears at random locations in the network in the absence of an external stimulus. Conversely, if recurrent excitation is too weak (via reduced E/I ratio), then the circuit cannot support persistent activity. In that regime, the persistent activity pattern collapses during the delay period and the memory is lost (upper left corner in Fig. 3). Of note, the loss of multistability that we observe, after only a moderate reduction of G_{EI} or G_{EE}, may result from the model's requirement that a single microcircuit be capable of both multistability and continuous representation of a memorandum. An extended model, which separates bistability and representation between different populations of cells, may exhibit greater robustness of multistability against disruption of E/I balance. However, the sensitivity of bump width to E/I ratio would still be present in such a configuration.

Disinhibition Increases Random Drift and Decreases Memory Precision

For robust WM, the precision of the encoded memorandum is critical and should be maintained throughout the delay. Stochastic activity internal to the neural system may interfere with maintenance, causing WM precision to decline continuously with increasing delay duration. In memory-guided saccade tasks, variability of saccade endpoints increases with delay duration (Funahashi et al., 1989; White et al. 1994; Ploner et al. 1998). Individuals with schizophrenia exhibit greater variability in memory-guided saccades, due in part to impaired maintenance (Badcock et al. 2008). Motivated by the observation that individuals with schizophrenia exhibit degraded precision during visuospatial WM maintenance, we characterized the effect of disinhibition on behavioral variability in the model. Specifically, we tested the time course of the variance of the WM report encoded in the network activity pattern. That is, we examined how the variability in WM report changes over time as a function of disinhibition, which would imply increased vulnerability to noise internal to the neural system.

Similar to experimental findings, the model exhibits a random drift of the angular location encoded by the WM bump. We extract the encoded angle, or report, from the population vector, which is the center of mass of the network activity pattern (Georgopoulos et al. 1986; see Materials and Methods). During the delay period, the encoded angle undergoes random drift, due to internal noise introduced by the Poisson background input. In essence, the network possesses a continuous family of possible bump states, so noise can shift the bump along this continuum. As the network integrates noise, the variability of report increases with delay duration (Compte et al. 2000; Chow and Coombes 2006; Carter and Wang 2007).

Figure 4*A* shows example traces of the encoded angle for control and disinhibition conditions. As expected, the variance of the encoded angle grows throughout the delay (Fig. 4*B*). We found that disinhibition increases the rate at which the variance grows (Fig. 4*B*,*E*). The increased variance is due to a deficit in maintenance rather than in initial encoding of the stimulus, since the variance at the start of the delay is similarly low for the 2 conditions. This effect of disinhibition can be understood through the effect of E/I ratio on bump width. Random drift occurs because the active neurons in the WM bump receive noisy background input. Pyramidal cells that are far from the bump are hyperpolarized by lateral inhibition and, therefore, farther from spike threshold than are cells that constitute the active flanks of the bump. As a result, background noise is more likely to produce output spikes in the depolarized cells on the flanks, consistent with an expansive nonlinearity in the firing rate versus mean current curve (e.g. Roxin et al. 2011). These noise-induced spikes then recurrently propagate to drive nearby cells and

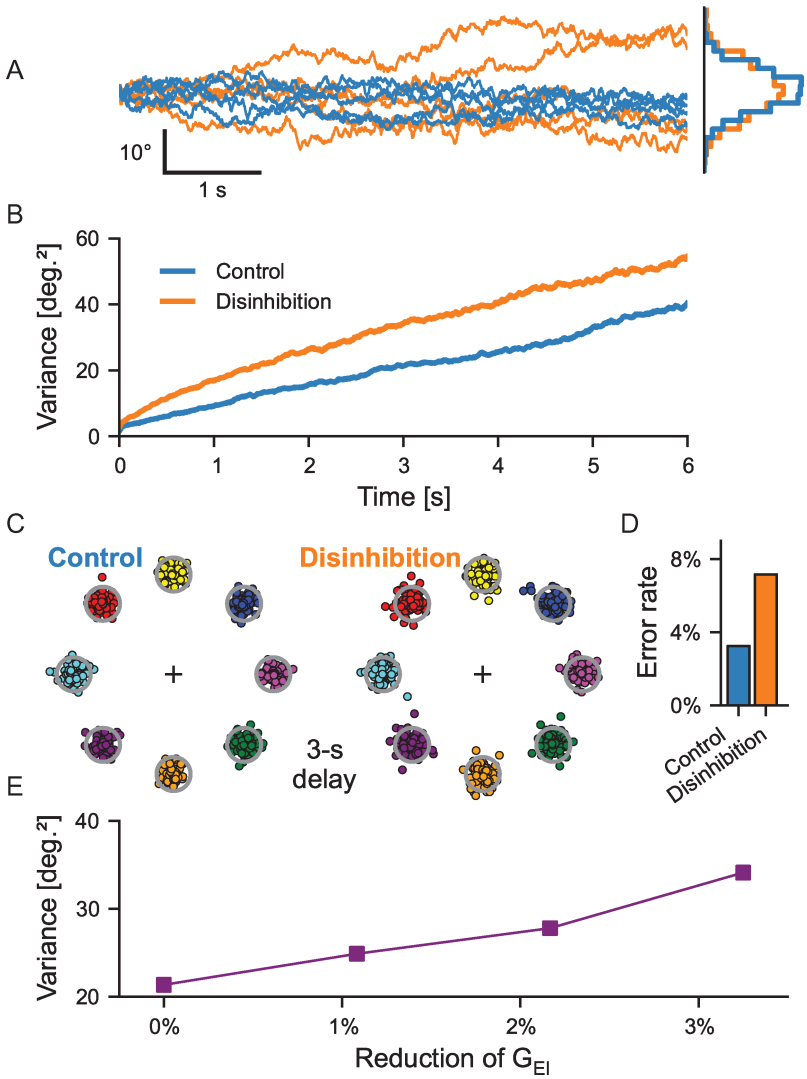

Figure 4. Disinhibition degrades memory precision due to increased noise-induced drift. (*A*) Decoding by the population vector of the WM activity pattern drifts across a 6-s delay. *Left:* Shown are 5 example traces each for control (blue) and disinhibition (orange). *Right:* Histogram of reports at the end of a 3-s delay. (*B*) The variance of the encoded angle fluctuations grows with time during the delay. Variance starts from a similar value for the 2 conditions, but the variance grows at a higher rate under disinhibition. (*C*) Visualization of reports after a 3-s delay for a memory-guided saccade task, with 96 trials at each of 8 stimulus angles. Gray circles mark the accuracy equivalent to within ±10° in the 1-dimensional model. (*D*) Error rates. Errors are counted if the reports shown in *C* lie outside of their corresponding gray circles. (*E*) Increase of the report variance, after a 3-s delay, as a function of the degree of disinhibition.

potentially to shift the bump's center of mass. A broadened bump under disinhibition is therefore subjected to more background noise input.

For visualization, we converted the variance of encoded angle into a 2D scatter of the memory report (Renart et al. 2003; see Materials and Methods), illustrating the degraded precision under disinhibition (Fig. 4*C*). If we introduce an accuracy threshold for correct responses (indicated with gray circles in Fig. 4*C*), the error rate is increased under disinhibition (Fig. 4*D*). This greater loss of WM precision over time captures prior experimental observations of behavior in schizophrenia (Badcock et al. 2008).

Disinhibition Increases the Range of Vulnerability to External Distractors

In addition to robust stimulus maintenance across a delay, a neural circuit that subserves WM should be resistant to intervening external distractors during the delay (Jonides et al. 2008). There is a rich literature suggesting that schizophrenia is associated with deficits in gating of external stimuli (Geyer et al. 2001; Turetsky et al. 2007) and abnormalities in resistance to distractors during WM (Hahn et al. 2010; Anticevic et al. 2011). In experimental settings, the effect of a distractor on WM appears to be largely dependent on the similarity between a distractor and the target held in memory. In visuospatial WM, similarity is determined by the spatial separation between the target and the distractor. Prior experimental studies in healthy human subjects found that a distractor can attract the memory report toward its location, but only if the distractor appears within a "distractibility window" around the target location (Van der Stigchel et al. 2007; Herwig et al. 2010), as predicted by the similarity hypothesis. In other words, WM report can be perturbed toward the distractor if target–distractor similarity is high enough (i.e. the distractor is near the target and overlaps the representation), but is unperturbed if target–distractor

similarity is low (i.e. the distractor is far away from the target and can be effectively filtered).

Previous work has shown that the present computational model possesses an inherent resistance to distractors (Compte et al. 2000; Gruber et al. 2006). Interestingly, model behavior also exhibits a dependence on target–distractor similarity. In that sense, there are 2 qualitative types of distractors, "near" and "far". When the distractor is spatially far from target, there is no overlap between the neural representations of the distractor and of the target in WM. In this regime, the interaction between the memory and distractor is "winner-take-all" competition mediated by lateral inhibition: If the distractor strength is low, the distractor will have negligible effect on the report, but if the distractor strength is too high, the memory bump will switch completely to the distractor location. In contrast, when the distractor is near the target, there is overlap between the distractor and target representations. If the distractor input excites neurons on the flanks of the bump, the target–distractor interaction is better described as "vector-averaging", and the resulting bump will lie at an intermediate location (Compte et al. 2000; Liu and Wang 2008). We define the model's distractibility window as the angular separation that results in the largest deviation in the report from the original cue location.

Since both experimental results in humans and model results exhibit a distractibility window within which distractors are able to attract the WM report, the model readily allows us to study the neural circuit basis of increased distractibility by disinhibition. In the model, the distractibility window is dependent on the width of the WM bump. As noted, attraction by the distractor occurs when there is overlap between the neural representations of the target in WM and of the distractor. Since disinhibition broadens WM representations, we specifically hypothesized that the distractibility window will be similarly expanded. This would increase the range of locations at which distractors can disrupt WM following disinhibition.

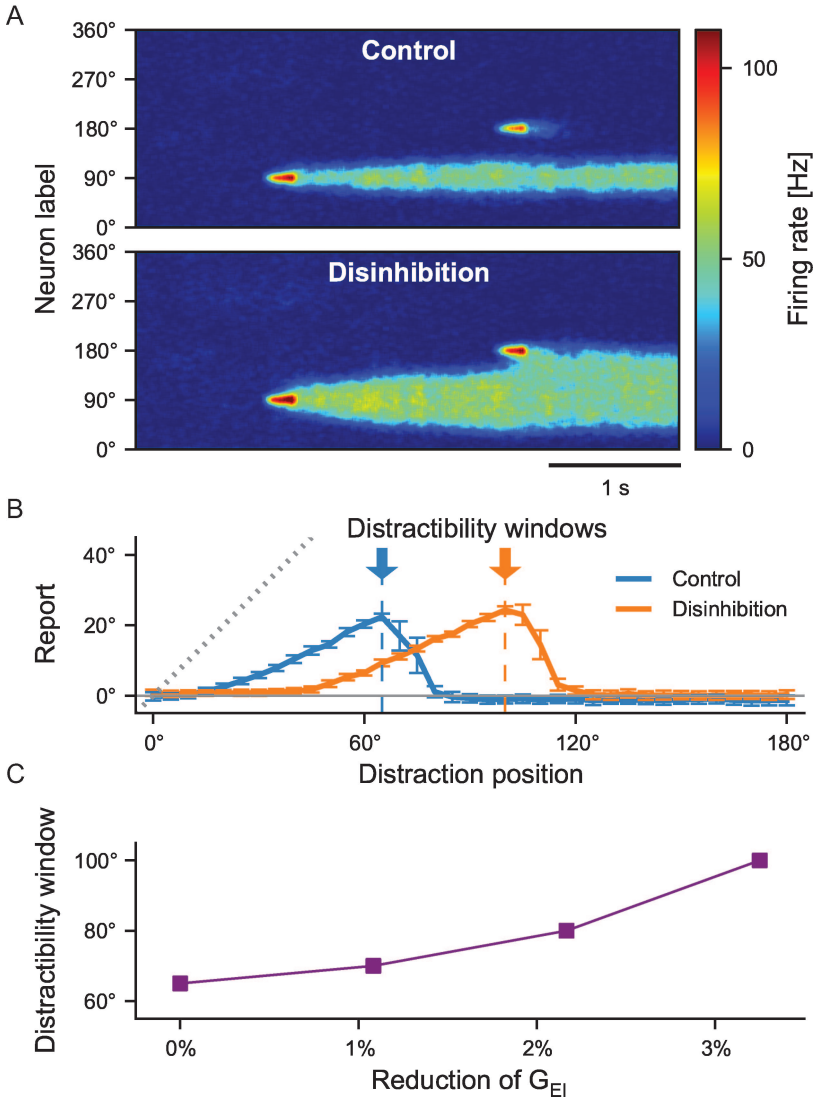

Figure 5. Disinhibition is detrimental the network's ability to filter out nonlocal distractors. (*A*) Spatiotemporal plot of network activity in response to a distractor presented during the delay at a distance of 90° from the target. The 250-ms distractor is presented at 1.5 s into the delay. (*B*) Deviation of report, read out after a 3-s delay, as a function of the angular distance between the distractor and the target. The angle corresponding to maximum displacement defines the distractibility window. The distractibility window is widened by disinhibition. Error bars show standard deviation. (*C*) Increase of the distractibility window as a function of the degree of disinhibition.

As predicted, disinhibition increases the width of the bump such that a far distractor is transformed to a near distractor (Fig. 5*A*). Put differently, the distractibility window increases under disinhibition (Fig. 5*B*,*C*), and disinhibition therefore widens the range of distractors that can disrupt WM. This model finding also suggests that manipulating target–distractor similarity (to measure the distractibility window) may be a highly sensitive experimental probe of changes in the neural WM representation (see Discussion).

In addition to widening the distractibility window, disinhibition changes the shape of the distractibility curve shown in Figure 5*B*. Interestingly, for an intermediate range within the distractibility window, disinhibition decreases the magnitude of deflection toward the distractor, which may be attributed to greater "inertia" of the broadened bump. The shapes of the distractibility curves in Figure 5*B* can be roughly decomposed into 4 regions: 1) very close (<20–30°) where neither network loses fidelity; 2) near (approximately 30–75°) where the control network is more distractible than the disinhibited network; 3) medium-range (approximately 75–120°) where the disinhibited network is affected by the distractor but the control network is not; and 4) distal (>120°) where neither network is affected by the distractor. We know of no relevant experimental studies that have combined 1) parametric control of target–distractor similarity; 2) direct, continuous report; and 3) direct comparison between control condition and a condition of putative disinhibition. The characterization of how disinhibition affects the similarity dependence of distractibility should be addressed by future psychophysical studies in healthy adults, clinical populations, and animal models.

Experimental Evidence for Broadened WM Representation

As described above, the core model prediction is that WM representations are broadened under disinhibition, due to disrupted E/I balance. We tested this prediction experimentally in humans using a pharmacological manipulation that is hypothesized to induce disinhibition (Anticevic et al. 2012). Specifically, we employed NMDAR antagonism via acute, low-dose administration of ketamine. As a pharmacological model of schizophrenia, ketamine allows for controlled and transient NMDAR manipulation in healthy volunteers (Krystal et al. 2003). Ketamine-induced neural disinhibition effects can be observed in both in vitro and in vivo (Greene 2001). While the precise pathways by which ketamine induces disinhibition are not fully resolved, at the low concentrations employed in human experimental work, there is evidence that its effects may involve a preferential blockade of NMDAR subunits (Kotermanski and Johnson 2009) that are expressed more in interneurons than in pyramidal cells (Xi et al. 2009).

In the experiment, healthy subjects performed a delayed spatial WM match/nonmatch task, following either saline (control) or ketamine infusion. In this task, subjects encoded multiple target locations, and after a delay period, indicated whether a probe was presented at a target location (match) or nontarget location (nonmatch; Fig. 6*A*). There are 2 types of errors that can occur in this task: Misses (nonmatch response to a probe at a target location), and false alarms (match response to a probe at a nontarget location).

In neural circuit models of match/nonmatch decision making, the probability of match response is a monotonically increasing function of the overlap between the neural representations of the probe and of the target in WM (Tagamets and Horwitz 1998; Engel and Wang 2011). For nontarget probes, the rate of false alarms increases with the overlap between the target memory and probe representations. Put simply, a match response to a nontarget probe is more likely when there is a larger overlap between the probe and the target memoranda.

Therefore, changes to the WM representation by disinhibition can potentially alter the target–probe overlap and thereby affect the pattern of responses. The model predicts that the increase in errors is selective, with an increased rate of false alarms when the probe is presented at near, but not far, nontarget locations. This is because the overlap between the target WM representation and the probe representation is increased substantially when the probe is presented at near nontarget locations (due to increased bump width), but not when presented at far nontarget locations. Similarly, the model predicts that misses should not increase since the overlap does not substantially decrease when the probe is presented at the target location. In other words, because the peak firing rate of the WM bump does not change substantially following disinhibition, there is similar overlap for a probe presented at the target location.

Figure 6*B* shows the effects of ketamine on error rates for false alarms to near nontarget probes, false alarms to far nontarget probes, and misses to target probes. To quantitatively test the model prediction using experimental data, we computed a repeated-measures ANOVA with the number of false alarms per subject as the dependent measure, and infusion (saline vs. ketamine), and distance (near vs. far nontarget probes) as within-subject factors. Experimental results revealed a significant infusion × distance interaction ($F_{1,19} = 7.55$, $P < 0.015$). As predicted, the interaction effect was driven by a significantly greater number of false alarms following ketamine infusion for near but not far distraction ($t_{(19)} = 4.09$, $P < 0.0007$, Cohen's $d = 1.1$), whereas this effect was substantially attenuated and did not reach significance for placebo trials ($t_{(19)} = 0.99$, $P = 0.33$, Cohen's $d = 0.25$). Moreover, when examining miss trials there was no significant difference across infusions ($t_{(19)} = 0.72$, $P = 0.35$). This selective increase in the number of errors for false alarms to near nontarget probes, but neither for false alarms to far nontarget probes nor for misses, is in line with the model prediction.

Moreover, as an explicit demonstration of the model prediction, we used disinhibition in the model to generate the selective pattern of errors we observed experimentally (Fig. 6*C*,*D*). The match/nonmatch decision is based on the overlap between the neural representations of the target in WM and of the probe (see Materials and Methods). The probability of match response is taken as a sigmoidal function of this target–probe overlap. As shown in Figure 6*D*, broadening of the WM bump by disinhibition is sufficient to capture the pattern of errors observed experimentally, with a selective increase in false alarms to near nontarget probes, but little change in false alarms to far nontarget probes or in misses. Of note, Figure 6B shows error rates for trials of each probe type for which subjects responded, to allow comparison to model performance (Figure 6*D*).

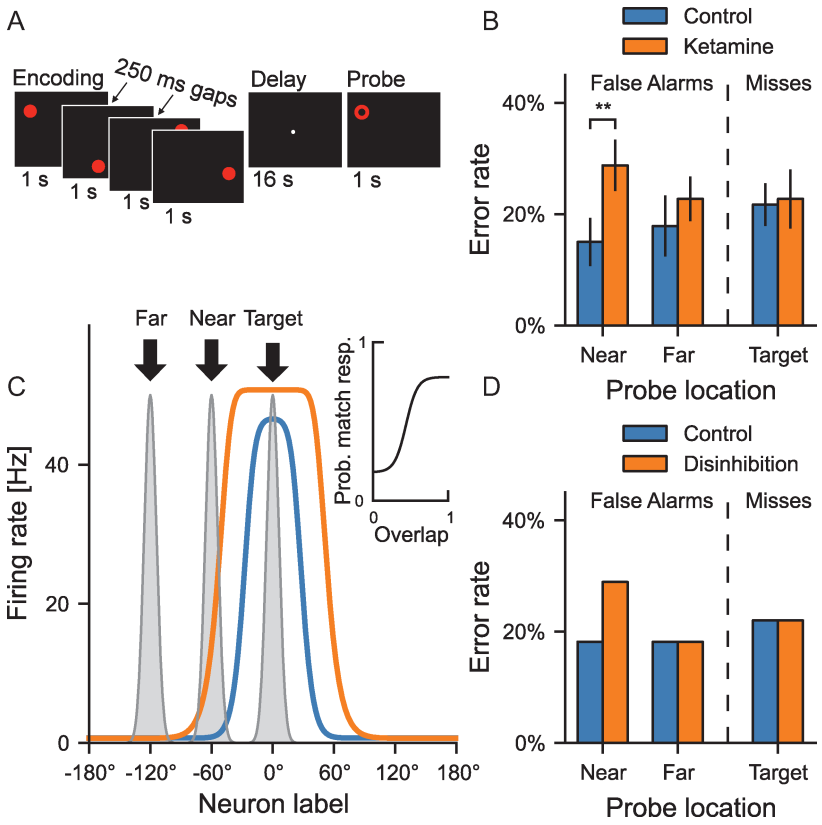

Figure 6. Ketamine induces errors in spatial WM specifically for near distractors, as predicted by the model. (*A*) Experimental task design. Subjects encoded spatial targets in WM, then after a delay, responded whether a probe matched a target location (match) or not (nonmatch). (*B*) Pattern of error rates for human subjects under placebo and ketamine. Ketamine degrades performance selectively, increasing false alarms to near distractor probes (**$P < 0.01$), but not false alarms to far distractor probes or misses to target probes (not significant). Here we computed error rates for each subject to allow comparison to model performance; comprehensive statistics in the main text pertain to original error counts across relevant factors of interest. (*C*) Illustration of how disinhibition in the WM model affects match/nonmatch decisions. Overlap between the probe representations (gray) and the target representation in WM is larger for disinhibition (orange) than for control (blue). Overlap increases most for the near distractor location. *Inset:* The probability of a match response is given by a sigmoidal function of the overlap. (*D*) Model performance under disinhibition replicates the observation of the human study that the error rate is differentially increased for false alarms to near distractor probes. As shown in *C*, we modeled probe locations as 60° and 120° from the target for near and far distractors, respectively.

Interestingly, a recent psychophysical study using a similar spatial WM match/nonmatch task found that, relative to healthy controls, individuals with schizophrenia also exhibit a selective increase in false alarms for near nontarget probes but not for far nontarget probes (Mayer and Park 2012). In addition to its effects on the maintenance of WM representations, NMDAR antagonism may also induce deficits downstream of WM, in the neural circuitry implementing the target–probe comparison and match/nonmatch decision. Future experimental and modeling studies should differentiate deficits due to WM from deficits due to decision making. Furthermore, a circuit model that incorporates the diversity of functional cell types in prefrontal cortex may be necessary to explain the diverse effects of NMDAR antagonists at both the neural and behavioral levels.

Restoring E/I Balance Ameliorates Deficits

The model suggests that E/I balance is the crucial parameter that determines WM performance. As illustrated in the parameter space characterization in Figure 3, if E/I balance is disrupted by a change at one site in the network, it can be restored through a compensatory change at another site (e.g. reduction in G_{EE} compensating for reduction in G_{EI}). Therefore, we explicitly tested the hypothesis that compensations that restore E/I balance in the model can ameliorate the observed WM deficits induced by disinhibition.

To this end, we implemented 2 compensations, modifying glutamatergic and GABAergic neurotransmission, respectively. Our first compensation is to decrease the strength of presynaptic glutamate release at recurrent synapses. Reducing presynaptic glutamate release has been proposed as compensation for the elevated E/I ratio and behavioral symptoms arising from schizophrenia (Large et al. 2005) and NMDAR antagonists (Moghaddam and Adams 1998; Anand et al. 2000; Brody et al. 2003; Krystal et al. 2005). Our second compensation is to increase the inhibitory conductance onto pyramidal cells (i.e. G_{IE}). This in effect strengthens the recruitment of inhibition, which is weakened by the disinhibition manipulation in the model. Lewis et al. (2004) proposed that this could be realized using activators of the $\alpha 2$ subunit of the GABAR (which mediates perisomatic inhibition to the axon initial segment of pyramidal cells), and an $\alpha 2$-subunit agonist was found to ameliorate WM deficits induced by ketamine in monkeys (Castner et al. 2010).

We examined whether these proof-of-principle compensations affected 4 key model properties outlined above: 1) baseline firing rate, 2) bump width, 3) drift due to internal noise, and 4) distractibility window in response to external

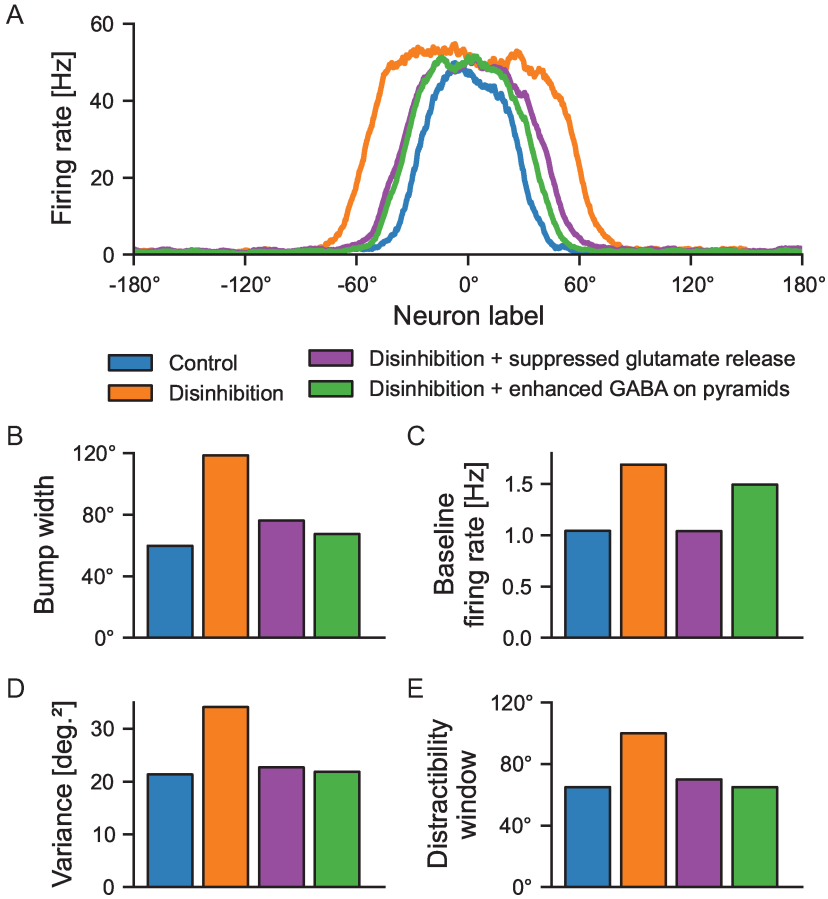

Figure 7. Compensations can restore E/I balance and ameliorate behavioral deficits. Disinhibition is combined with either suppression of presynaptic glutamate release at recurrent synapses (purple) or enhanced GABAR conductance on pyramidal cells (green). (*A*) Bell-shaped persistent activity profile (WM bump) for control, disinhibition, and the 2 compensation conditions. (*B*) The compensations restore the bump width from its broadened value under disinhibition to near its control value. (*C*) The disinhibition of the spontaneous firing rate is ameliorated by the compensations, particularly the glutamatergic manipulation. (*D*) The report variance, after a 3-s delay, is reduced to near the control level by the compensations. (*E*) The distractibility window is reduced to near the control level by the compensations.

distractors. Implemented in conjunction with our disinhibition manipulation, both compensations reduced the width of the bump attractor (Fig. 7*A*,*B*). We set the strength of the compensation such that bump width was restored to near its control value, and then characterized neural and behavioral effects.

In addition to restoring bump width, both compensations reduce the baseline firing rates from their values under disinhibition (Fig. 7*C*). Both compensations therefore ameliorate the neural changes induced by disinhibition, both at baseline and during active WM maintenance. We observed that the decrease in baseline firing was smaller for the GABAergic compensation, indicating that strengthening GABAR conductance on pyramidal cells is more effective at reducing bump width than at reducing baseline firing rates in the model.

At the behavioral level, both compensations ameliorated WM deficits observed under disinhibition. Report variance and distractibility windows reduced toward their control values (Fig. 7*D*,*E*). These manipulations are a simple proof-of-principle, rather than detailed models of any given experimental or actual treatment manipulation. However, that these 2 manipulations restored behavior, despite different sites of action, highlights that the critical parameter of the network is E/I ratio, which controls bump width.

It is also important to note the magnitudes of the 2 compensations. We implemented a 25% decrease in presynaptic glutamate release for the first compensation, but only a 2% increase in GABAR conductance on pyramidal cells for the second compensation. This order-of-magnitude difference reflects that the glutamatergic compensation in the model acted both on pyramidal cells (decreasing E/I ratio) and on interneurons (increasing E/I ratio). Because these 2 changes partially counteract one another, the level of the manipulation must be relatively large to restore E/I balance. In contrast, because the GABAergic compensation in the model acted only on pyramidal cells, a relatively lower level of manipulation is required to restore E/I balance. Together, this suggests that compensations with greater specificity, with strong changes in either pyramidal cells or interneurons but not both, may be more efficacious than those that act equally on pyramidal cells and interneurons.

Discussion

In this study, we examined the neural and behavioral effects of disinhibition in a cortical microcircuit model of WM. We found that recurrent circuit function is highly sensitive to perturbations of E/I balance; a slight weakening of inhibition

results in marked neural and behavioral changes. The salient neural changes are elevated E/I ratio and broadened WM representation. They are linked to behavioral deficits, namely increased susceptibility to both internal noise and external distraction. Our proof-of-principle compensations further highlight that E/I balance is a critical property of the model that relates to observed deficits induced by disinhibition.

This work focuses on the consequences of a reduced overall strength of connections from excitatory cells to interneurons, which could be related, at least in part, to interneuron dysfunction hypothesized to operate in schizophrenia (Lewis et al. 2005; Lisman et al. 2008; Marin 2012). Previous modeling studies have suggested that NMDA receptors on pyramidal cells may be important for providing sufficiently strong recurrent synaptic excitation that underlies persistent activity during WM (Wang 1999; Durstewitz et al. 2000). The present work provides a different, complementary perspective that emphasizes the importance of the width of the population firing pattern, rather than the peak firing rate. The width of the WM bump is controlled by E/I balance. Increasing E/I ratio, and thereby bump width, degrades WM robustness in a manner that yields specific types of errors. Such errors are dissociable from those predicted by a complete loss of WM maintenance, which may result from a collapse of persistent activity due to insufficient recurrent excitation.

Implications for Task Design

Present findings offer important implications for the paradigm design of WM tasks used to study cognitive dysfunction in neuropsychiatric disease. First, the present model suggests that a continuous, veridical measure of the report is sensitive to detecting degraded precision due to drift. This increase in scatter may be subtle and difficult to detect with a categorical (e.g. match/nonmatch) design. The veridical report of memoranda can also help to distinguish deficits in maintenance from potential deficits in the downstream target–probe comparison, and match/nonmatch decision computations inherent in the match/nonmatch paradigm (Lencz et al. 2003; Zhang and Luck 2008). Building on this continuous measure of report, future studies can vary delay duration to better distinguish increased delay-dependent drift during maintenance from impaired encoding (Lee and Park 2005; Wei et al. 2012).

Our findings also underscore that target–distractor similarity is a critical parameter in determining the effect of distractors on WM (Van der Stigchel et al. 2007; Herwig et al. 2010). To characterize how this changes under disinhibition, future studies should employ stimulus sets with well-defined target–distractor similarity and manipulate target–distractor similarity parametrically. The veridical report is critical in this setting, in order to measure how distractors alter the memory of the target. We find that the distractibility window varies with bump width as E/I balance is perturbed. This effect suggests an interesting interpretation of behavior in distractor tasks, which extends beyond their use to characterize how dysfunction changes distractibility. That is, distractors can serve as a sensitive probe of the underlying neural WM representation, enabling one to measure behaviorally how neural dysfunction perturbs WM representation.

Moreover, the simple tasks that we modeled (i.e. single-item spatial WM with continuous report and parametric control of target–distractor similarity) explicitly build upon electrophysiological studies in awake behaving monkeys (Funahashi et al. 1989). These tasks can therefore be carried out in nonhuman primates as well as clinical populations and healthy volunteers undergoing pharmacological manipulations. This framework in turn allows for the complete translation from the model to electrophysiology, pharmacology, and ultimately clinical phenomena. Our proof-of-principle demonstration of compensations for disinhibition points toward future computational studies of the neural and behavioral effects of pharmacological treatments.

Relation to Other Modeling Studies

Present findings complement recent studies using attractor-based WM models to link the synaptic changes to behavioral deficits in schizophrenia (Loh et al. 2007; Cano-Colino and Compte 2012). Both of these previous computational studies focused on very strong microcircuit perturbations, to such an extent that they compromise network multistability. In contrast, our study focused on more subtle effects that occur within the multistability regime (Fig. 3). Both prior studies describe a more extreme disinhibition regime in which the baseline state is destabilized, and a WM pattern at a random location ("false memory") can spontaneously emerge anywhere in the network without a sensory cue. In addition, the discrete-pool model used in Loh et al. (2007) cannot capture features that emerge naturally from a continuous model described here: Broadening of WM representations, smooth random drift of memory, and sensitivity to target–distractor similarity. Detecting all of these effects is dependent on the continuous nature of the present model.

As noted, because these previous studies put their networks in regimes of compromised multistability, they predict the different patterns of degraded behavior and errors from those predicted by the present subtle perturbation—namely, they predict the random formation of erroneous WM patterns (false memory). In the context of a match/nonmatch task, spontaneous bump creation at random locations would increase false alarms to far nontarget probes. Increased miss rate would be expected if the persistent activity pattern was completely lost in the target cells (Loh et al. 2007), or a bump spontaneously appeared in nontarget cells, interfering with target encoding (Cano-Colino and Compte 2012). Neither of these effects, the increased false alarm rate to far nontarget probes or the increased miss rate, were observed in the reported ketamine experiment (Fig. 6*C*). It is worth noting that the same model (Compte et al. 2000) was used by Cano-Colino and Compte (2012) and in the present work. In that sense, even relatively modest changes of parameters, such as the degree of disinhibition, can give rise to qualitatively different behavioral effects with vastly different clinical implications (namely loss of internal precision vs. false memory formation).

Indeed, in a single-item spatial WM experiment, behavioral results for schizophrenia patients were consistent with degraded precision, manifested as increased scatter of reports around the target location (Badcock et al. 2008). This behavioral pattern and those reported by other studies (Mayer and Park 2012) appear inconsistent either with the complete loss of the memory or with false memories, since both of these regimes would correspond with reports made to random, far locations. However, given schizophrenia population heterogeneity, it is possible that some patients indeed show patterns

consistent with more extreme forms of impairment. Therefore, it will be vital for future studies to examine whether any experimental conditions match the more extreme forms of disinhibition regimes proposed in previous modeling studies (Loh et al. 2007; Cano-Colino and Compte 2012). Together, these differing model predictions underscore the utility of distinguishing different error types. The detailed characterization of behavior reports and error types can be harnessed to constrain which network regimes in models may correspond to experimentally studied disease states.

Future Directions

In this work, we focused on single-item spatial WM, since it has a better understood neural circuit basis compared with other forms of WM (Funahashi et al. 1989). There is evidence that WM capacity is impaired in schizophrenia (Gold et al. 2010). Computational models suggest that the maintenance of multiple items within a WM circuit depends on the strength of lateral inhibition (Edin et al. 2009) and on the ability to keep activity patterns separated from each other (Wei et al. 2012). Future work with those models could implement disinhibition and test whether elevated E/I ratio, with broadened representations, reduces WM capacity.

Elevated E/I ratio has been implicated in the neuropathology of other psychiatric disorders, including autism spectrum disorders (Rubenstein and Merzenich 2003; Yizhar et al. 2011) and attention-deficit/hyperactivity disorder (Won et al. 2011), both of which exhibit WM deficits (Martinussen et al. 2005; Williams et al. 2005). Present findings on cortical disinhibition may generally describe a core neuropathology underlying multiple neuropsychiatric disorders. Sensitive psychophysical characterization of WM deficits across clinical populations could clarify similarities and differences in WM behavior and potentially elucidate underlying neural substrates of dysfunctions.

Elevated prefrontal E/I ratio may result from alterations in multiple pathways that are implicated in schizophrenia. These include GABAergic deficits, such as reduced GAD67 production (Lewis et al. 2005); increased background input from other areas such as the thalamus or hippocampus (Lisman et al. 2008); or abnormal levels of neuromodulators (Laruelle et al. 2003). We predict that multiple mechanisms that elevate E/I ratio may converge in their neural and behavioral consequences. Similarly, there may be multiple mechanisms to restore E/I balance, through compensations in the glutamatergic, GABAergic, or neuromodulatory systems. Computational models linking synaptic perturbations to behavior provide a powerful platform to explore network sensitivity to compensation by various interventions.

As a future extension, the study of disinhibition in recurrent networks should incorporate the diversity of inhibition observed in the cortex. In particular, 2 important classes of interneurons are those that target the perisomatic region of pyramidal cells, and those that target distal dendrites of pyramidal cells (Markram et al. 2004). This segregation of targets may allow these interneuron classes to serve distinct roles in cortical computation (Lovett-Barron et al. 2012; Royer et al. 2012). Both interneuron cell classes are altered in schizophrenia (Lewis et al. 2005; Morris et al. 2008; Bitanihirwe et al. 2009). Interestingly, these interneuron classes and pyramidal cells may be differentially sensitive to NMDAR antagonists (Neymotin et al. 2011) and express different proportions of AMPARs and subunits of NMDARs across developmental stages (Wang and Gao 2009; Xi et al. 2009; Rotaru et al. 2011), potentially contributing to the complex developmental trajectory of schizophrenia. Circuit models that incorporate a division of labor among different classes of interneurons (Wang et al. 2004) may be able to link the disruption of each class to specific behavioral deficits. Since soma- and dendrite-targeting interneurons are differentially engaged in gamma- and theta-band oscillations (Gonzalez-Burgos and Lewis 2008; Wang 2010), more detailed microcircuit models have the potential to link the specific aspects of cognitive dysfunction to the diversity of abnormal cortical oscillations observed in schizophrenia (Uhlhaas and Singer 2010).

Funding

This work was supported by the National Institutes of Health (T15-LM07056 to J.D.M., DP5-OD012109-01 to A.A., and R01-MH062349 to X.-J.W.) AstraZeneca Pharmaceuticals.

Notes

We thank Dr Naomi R. Driesen for her assistance with experimental paradigm development. *Conflict of Interest*: J.H.K. is a consultant to the companies listed below which cost less than $10,000 per year: Aisling Capital, LLC Astellas Pharma Global Development, Inc., AstraZeneca Pharmaceuticals, Biocortech, Brintnall & Nicolini, Inc., Easton Associates, Gilead Sciences, Inc., GlaxoSmithKline, Janssen Pharmaceuticals, Lundbeck Research USA, Medivation, Inc., Merz Pharmaceuticals, MK Medical Communications, F. Hoffmann-La Roche Ltd., Sage Therapeutics, Inc., SK Holdings Co., Ltd., Sunovion Pharmaceuticals, Inc., Takeda Industries, Teva Pharmaceutical Industries, Ltd. and also member of Scientific Advisory Board to Abbott Laboratories, Bristol-Myers Squibb, CHDI Foundation, Inc., Eisai, Inc., Eli Lilly and Co., Forest Laboratories, Inc., Lohocla Research Corporation, Mnemosyne Pharmaceuticals, Inc., Naurex, Inc., Pfizer Pharmaceuticals, Shire Pharmaceuticals, StratNeuro Research Program at Karolinska Institute (International Advisory Board). He is one of the Board of Directors to Coalition for Translational Research in Alcohol and Substance Use Disorders and President of American College of Neuropsychopharmacology. He is Editor (income greater than $10,000) of the journal *Biological Psychiatry*. He has patents and inventions on: 1) Seibyl JP, Krystal JH, Charney DS. Dopamine and noradrenergic reuptake inhibitors in treatment of schizophrenia. Patent #:5,447,948. September 5, 1995; 2) he is a co-inventor with Dr. Gerard Sanacora on a filed patent application by Yale University related to targeting the glutamatergic system for the treatment of neuropsychiatric disorders (PCTWO06108055A1); and 3) Intranasal Administration of Ketamine to Treat Depression (pending).

References

Amit DJ, Brunel N. 1997. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cereb Cortex. 7:237–252.

Anand A, Charney DS, Oren DA, Berman RM, Hu XS, Cappiello A, Krystal JH. 2000. Attenuation of the neuropsychiatric effects of ketamine with lamotrigine: support for hyperglutamatergic effects of *N*-methyl-D-aspartate receptor antagonists. Arch Gen Psychiatry. 57:270–276.

Anticevic A, Gancsos M, Murray JD, Repovs G, Driesen NR, Ennis DJ, Niciu MJ, Morgan PT, Surti TS, Bloch MH et al. 2012. NMDA receptor function in large-scale anticorrelated neural systems with implications for cognition and schizophrenia. Proc Natl Acad Sci USA. 109:16720–16725.

Anticevic A, Repovs G, Corlett PR, Barch DM. 2011. Negative and nonemotional interference with visual working memory in schizophrenia. Biol Psychiatry. 70:1159–1168.
Badcock JC, Badcock DR, Read C, Jablensky A. 2008. Examining encoding imprecision in spatial working memory in schizophrenia. Schizophr Res. 100:144–152.
Barch DM, Ceaser A. 2012. Cognition in schizophrenia: core psychological and neural mechanisms. Trends Cogn Sci. 16:27–34.
Bitanihirwe BK, Lim MP, Kelley JF, Kaneko T, Woo TU. 2009. Glutamatergic deficits and parvalbumin-containing inhibitory neurons in the prefrontal cortex in schizophrenia. BMC Psychiatry. 9:71.
Brody SA, Geyer MA, Large CH. 2003. Lamotrigine prevents ketamine but not amphetamine-induced deficits in prepulse inhibition in mice. Psychopharmacology (Berl). 169:240–246.
Brunel N, Wang X-J. 2001. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. J Comput Neurosci. 11:63–85.
Cano-Colino M, Compte A. 2012. A computational model for spatial working memory deficits in schizophrenia. Pharmacopsychiatry. 45(Suppl 1):49–56.
Carter E, Wang X-J. 2007. Cannabinoid-mediated disinhibition and working memory: dynamical interplay of multiple feedback mechanisms in a continuous attractor model of prefrontal cortex. Cereb Cortex. 17(Suppl 1):16–26.
Castner SA, Arriza JL, Roberts JC, Mrzljak L, Christian EP, Williams GV. 2010. Reversal of ketamine-induced working memory impairments by the GABAA $\alpha 2/3$ agonist TPA023. Biol Psychiatry. 67:998–1001.
Chow C, Coombes S. 2006. Existence and wandering of bumps in a spiking neural network model. SIAM J Appl Dyn Syst. 5:552–574.
Compte A, Brunel N, Goldman-Rakic PS, Wang X-J. 2000. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb Cortex. 10:910–923.
Domino SE, Domino LE, Domino EF. 1982. Comparison of two and three compartment models of phencyclidine in man. Subst Alcohol Actions Misuse. 3:205–211.
Driesen NR, Leung HC, Calhoun VD, Constable RT, Gueorguieva R, Hoffman R, Skudlarski P, Goldman-Rakic PS, Krystal JH. 2008. Impairment of working memory maintenance and response in schizophrenia: functional magnetic resonance imaging evidence. Biol Psychiatry. 64:1026–1034.
Durstewitz D, Seamans JK, Sejnowski TJ. 2000. Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. J Neurophysiol. 83:1733–1750.
Edin F, Klingberg T, Johansson P, McNab F, Tegner J, Compte A. 2009. Mechanism for top-down control of working memory capacity. Proc Natl Acad Sci U S A. 106:6802–6807.
Elvevåg B, Goldberg T. 2000. Cognitive impairment in schizophrenia is the core of the disorder. Crit Rev Neurobiol. 14:1–21.
Engel TA, Wang X-J. 2011. Same or different? A neural circuit mechanism of similarity-based pattern match decision making. J Neurosci. 31:6982–6996.
Funahashi S, Bruce CJ, Goldman-Rakic PS. 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol. 61:331–349.
Fuster JM. 2008. The prefrontal cortex. 4th ed. New York: Academic Press.
Georgopoulos AP, Schwartz AB, Kettner RE. 1986. Neuronal population coding of movement direction. Science. 233:1416–1419.
Geyer MA, Krebs-Thomson K, Braff DL, Swerdlow NR. 2001. Pharmacological studies of prepulse inhibition models of sensorimotor gating deficits in schizophrenia: a decade in review. Psychopharmacology (Berl). 156:117–154.
Gold JM, Hahn B, Zhang WW, Robinson BM, Kappenman ES, Beck VM, Luck SJ. 2010. Reduced capacity but spared precision and maintenance of working memory representations in schizophrenia. Arch Gen Psychiatry. 67:570–577.
Goldman-Rakic PS. 1994. Working memory dysfunction in schizophrenia. J Neuropsychiatry Clin Neurosci. 6:348–357.
Gonzalez-Burgos G, Lewis DA. 2008. GABA neurons and the mechanisms of network oscillations: implications for understanding cortical dysfunction in schizophrenia. Schizophr Bull. 34:944–961.
Goodman D, Brette R. 2008. Brian: a simulator for spiking neural networks in python. Front Neuroinform. 2:5.
Greene R. 2001. Circuit analysis of NMDAR hypofunction in the hippocampus, in vitro, and psychosis of schizophrenia. Hippocampus. 11:569–577.
Gruber AJ, Dayan P, Gutkin BS, Solla SA. 2006. Dopamine modulation in the basal ganglia locks the gate to working memory. J Comput Neurosci. 20:153–166.
Hahn B, Robinson BM, Kaiser ST, Harvey AN, Beck VM, Leonard CJ, Kappenman ES, Luck SJ, Gold JM. 2010. Failure of schizophrenia patients to overcome salient distractors during working memory encoding. Biol Psychiatry. 68:603–609.
Herwig A, Beisert M, Schneider WX. 2010. On the spatial interaction of visual working memory and attention: evidence for a global effect from memory-guided saccades. J Vis. 10:8.
Howell D. 2012. Statistical methods for psychology. 8th ed. Belmont, CA: Wadsworth.
Jahr CE, Stevens CF. 1990. Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. J Neurosci. 10:3178–3182.
Jonides J, Lewis RL, Nee DE, Lustig CA, Berman MG, Moore KS. 2008. The mind and brain of short-term memory. Annu Rev Psychol. 59:193–224.
Kotermanski SE, Johnson JW. 2009. Mg^{2+} imparts NMDA receptor subtype selectivity to the Alzheimer's drug memantine. J Neurosci. 29:2774–2779.
Krystal JH, Abi-Saab W, Perry E, D'Souza DC, Liu N, Gueorguieva R, McDougall L, Hunsberger T, Belger A, Levine L et al. 2005. Preliminary evidence of attenuation of the disruptive effects of the NMDA glutamate receptor antagonist, ketamine, on working memory by pretreatment with the group II metabotropic glutamate receptor agonist, LY354740, in healthy human subjects. Psychopharmacology (Berl). 179:303–309.
Krystal JH, D'Souza DC, Mathalon D, Perry E, Belger A, Hoffman R. 2003. NMDA receptor antagonist effects, cortical glutamatergic function, and schizophrenia: toward a paradigm shift in medication development. Psychopharmacology (Berl). 169: 215–233.
Krystal JH, Karper LP, Seibyl JP, Freeman GK, Delaney R, Bremner JD, Heninger GR, Bowers MB, Charney DS. 1994. Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. Arch Gen Psychiatry. 51:199–214.
Large CH, Webster EL, Goff DC. 2005. The potential role of lamotrigine in schizophrenia. Psychopharmacology (Berl). 181:415–436.
Laruelle M, Kegeles LS, Abi-Dargham A. 2003. Glutamate, dopamine, and schizophrenia: from pathophysiology to treatment. Ann N Y Acad Sci. 1003:138–158.
Lee J, Park S. 2005. Working memory impairments in schizophrenia: a meta-analysis. J Abnorm Psychol. 114:599–611.
Lencz T, Bilder RM, Turkel E, Goldman RS, Robinson D, Kane JM, Lieberman JA. 2003. Impairments in perceptual competency and maintenance on a visual delayed match-to-sample test in first-episode schizophrenia. Arch Gen Psychiatry. 60:238–243.
Leung HC, Gore JC, Goldman-Rakic PS. 2002. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. J Cogn Neurosci. 14:659–671.
Lewis DA, Hashimoto T, Volk DW. 2005. Cortical inhibitory neurons and schizophrenia. Nat Rev Neurosci. 6:312–324.
Lewis DA, Volk DW, Hashimoto T. 2004. Selective alterations in prefrontal cortical GABA neurotransmission in schizophrenia: a novel target for the treatment of working memory dysfunction. Psychopharmacology (Berl). 174:143–150.
Lisman JE, Coyle JT, Green RW, Javitt DC, Benes FM, Heckers S, Grace AA. 2008. Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. Trends Neurosci. 31:234–242.

Liu F, Wang X-J. 2008. A common cortical circuit mechanism for perceptual categorical discrimination and veridical judgment. PLoS Comput Biol. 4:e1000253.
Loh M, Rolls ET, Deco G. 2007. A dynamical systems hypothesis of schizophrenia. PLoS Comput Biol. 3:e228.
Lovett-Barron M, Turi GF, Kaifosh P, Lee PH, Bolze F, Sun XH, Nicoud JF, Zemelman BV, Sternson SM, Losonczy A. 2012. Regulation of neuronal input transformations by tunable dendritic inhibition. Nat Neurosci. 15:423–430.
Marin O. 2012. Interneuron dysfunction in psychiatric disorders. Nat Rev Neurosci. 13:107–120.
Markram H, Toledo-Rodriguez M, Wang Y, Gupta A, Silberberg G, Wu C. 2004. Interneurons of the neocortical inhibitory system. Nat Rev Neurosci. 5:793–807.
Martinussen R, Hayden J, Hogg-Johnson S, Tannock R. 2005. A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 44:377–384.
Mayer JS, Park S. 2012. Working memory encoding and false memory in schizophrenia and bipolar disorder in a spatial delayed response task. J Abnorm Psychol. 121:784–794.
Moghaddam B, Adams BW. 1998. Reversal of phencyclidine effects by a group II metabotropic glutamate receptor agonist in rats. Science. 281:1349–1352.
Morris HM, Hashimoto T, Lewis DA. 2008. Alterations in somatostatin mRNA expression in the dorsolateral prefrontal cortex of subjects with schizophrenia or schizoaffective disorder. Cereb Cortex. 18:1575–1587.
Nakazawa K, Zsiros V, Jiang Z, Nakao K, Kolata S, Zhang S, Belforte JE. 2012. GABAergic interneuron origin of schizophrenia pathophysiology. Neuropharmacology. 62:1574–1583.
Neymotin SA, Lazarewicz MT, Sherif M, Contreras D, Finkel LH, Lytton WW. 2011. Ketamine disrupts modulation of in a computer model of hippocampus. J Neurosci. 31:11733–11743.
Owen AM, McMillan KM, Laird AR, Bullmore E. 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp. 25:46–59.
Patil ST, Zhang L, Martenyi F, Lowe SL, Jackson KA, Andreev BV, Avedisova AS, Bardenstein LM, Gurovich IY, Morozova MA et al. 2007. Activation of mGlu2/3 receptors as a new approach to treat schizophrenia: a randomized phase 2 clinical trial. Nat Med. 13:1102–1107.
Ploner CJ, Gaymard B, Rivaud S, Agid Y, Pierrot-Deseilligny C. 1998. Temporal limits of spatial working memory in humans. Eur J Neurosci. 10:794–797.
Rao SG, Williams GV, Goldman-Rakic PS. 2000. Destruction and creation of spatial tuning by disinhibition: GABA(A) blockade of prefrontal cortical neurons engaged by working memory. J Neurosci. 20:485–494.
Renart A, Song P, Wang X-J. 2003. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. Neuron. 38:473–485.
Rotaru DC, Yoshino H, Lewis DA, Ermentrout GB, Gonzalez-Burgos G. 2011. Glutamate receptor subtypes mediating synaptic activation of prefrontal cortex neurons: relevance for schizophrenia. J Neurosci. 31:142–156.
Roxin A, Brunel N, Hansel D, Mongillo G, van Vreeswijk C. 2011. On the distribution of firing rates in networks of cortical neurons. J Neurosci. 31:16217–16226.
Royer S, Zemelman BV, Losonczy A, Kim J, Chance F, Magee JC, Buzsaki G. 2012. Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. Nat Neurosci. 15:769–775.
Rubenstein JL, Merzenich MM. 2003. Model of autism: increased ratio of excitation/inhibition in key neural systems. Genes Brain Behav. 2:255–267.
Shadlen MN, Newsome WT. 1994. Noise, neural codes and cortical organization. Curr Opin Neurobiol. 4:569–579.
Shu Y, Hasenstaub A, McCormick DA. 2003. Turning on and off recurrent balanced cortical activity. Nature. 423:288–293.
Tagamets MA, Horwitz B. 1998. Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. Cereb Cortex. 8:310–320.
Tsodyks M, Skaggs W, Sejnowski T, McNaughton B. 1997. Paradoxical effects of external modulation of inhibitory interneurons. J Neurosci. 17:4382–4388.
Tuckwell HC. 1988. Introduction to theoretical neurobiology. Cambridge: Cambridge University Press.
Turetsky BI, Calkins ME, Light GA, Olincy A, Radant AD, Swerdlow NR. 2007. Neurophysiological endophenotypes of schizophrenia: the viability of selected candidate measures. Schizophr Bull. 33:69–94.
Uhlhaas PJ, Singer W. 2010. Abnormal neural oscillations and synchrony in schizophrenia. Nat Rev Neurosci. 11:100–113.
Van der Stigchel S, Merten H, Meeter M, Theeuwes J. 2007. The effects of a task-irrelevant visual event on spatial working memory. Psychon Bull Rev. 14:1066–1071.
van Vreeswijk C, Sompolinsky H. 1996. Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science. 274:1724–1726.
Wang HX, Gao WJ. 2009. Cell type-specific development of NMDA receptors in the interneurons of rat prefrontal cortex. Neuropsychopharmacology. 34:2028–2040.
Wang X-J. 2010. Neurophysiological and computational principles of cortical rhythms in cognition. Physiol Rev. 90:1195–1268.
Wang X-J. 1999. Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. J Neurosci. 19:9587–9603.
Wang X-J, Tegnér J, Constantinidis C, Goldman-Rakic PS. 2004. Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory. Proc Natl Acad Sci USA. 101:1368–1373.
Wei Z, Wang X-J, Wang D-H. 2012. From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. J Neurosci. 32:11228–11240.
White JM, Sparks DL, Stanford TR. 1994. Saccades to remembered target locations: and analysis of systematic and variable errors. Vision Res. 34:79–92.
Williams DL, Goldstein G, Carpenter PA, Minshew NJ. 2005. Verbal and spatial working memory in autism. J Autism Dev Disord. 35:747–756.
Won H, Mah W, Kim E, Kim JW, Hahm EK, Kim MH, Cho S, Kim J, Jang H, Cho SC et al. 2011. GIT1 is associated with ADHD in humans and ADHD-like behaviors in mice. Nat Med. 17:566–572.
Xi D, Zhang W, Wang HX, Stradtman GG, Gao WJ. 2009. Dizocilpine (MK-801) induces distinct changes of N-methyl-D-aspartic acid receptor subunits in parvalbumin-containing interneurons in young adult rat prefrontal cortex. Int J Neuropsychopharmacol. 12:1395–1408.
Yizhar O, Fenno LE, Prigge M, Schneider F, Davidson TJ, O'Shea DJ, Sohal VS, Goshen I, Finkelstein J, Paz JT et al. 2011. Neocortical excitation/inhibition balance in information processing and social dysfunction. Nature. 477:171–178.
Zhang W, Luck SJ. 2008. Discrete fixed-resolution representations in visual working memory. Nature. 453:233–235.