Cognitive Control Errors in Nonhuman Primates Resembling Those in Schizophrenia Reflect Opposing Effects of NMDA Receptor Blockade on Causal Interactions Between Cells and Circuits in Prefrontal and Parietal Cortex

Supplemental Information

Supplemental Methods

Neuronal recording

We recorded neural activity from the dorsolateral prefrontal cortex (Brodmann area 46) and posterior parietal cortex (area 7a) concurrently using two computer driven 16-microelectrode drives (Thomas Recording GmbH). Action potentials of single neurons were isolated using a template-matching algorithm (Alpha Omega Engineering, Alpharetta, GA). Local field potentials (LFP) were recorded concurrently (sampled at 2000 Hz) on up to four of the same electrodes detecting spiking activity in each cortical area, band-pass filtered (1-100 Hz), notch filtered (60 Hz), demeaned, and resampled (1000 Hz) using the ft_preprocessing function in the FieldTrip open source Matlab toolbox (1). Channels or trials of LFP data were rejected if they contained movement artifacts or strong residual 60 Hz noise. All ensembles included in the analysis at least 200 trials of neural data.

NMDAR antagonist administration and neural database

At the start of each day's experiment, monkeys received either no injection, a systemic injection of saline, or a systemic injection of the NMDAR antagonist phencyclidine. In our earlier behavioral experiments investigating the influence of NMDAR blockade in nonhuman primates on cognitive control in the DPX task, we used the relatively short-acting NMDAR antagonist ketamine (2). However, given the necessity to obtain as much neural data as possible during the period of NMDAR antagonist-induced cognitive impairment, we selected the NMDAR antagonist phencyclidine for the current study given its longer duration of action. Collecting more neural data from each ensemble in the Drug condition allowed us to more accurately estimate conditional probabilities between neural signals improving the robustness of the causal discovery analysis. After the injection, monkeys were placed in the behavioral testing apparatus. We then independently advanced 16 thin glass-coated microelectrodes through the dura and into the prefrontal cortex, and an additional 16 electrodes into the posterior parietal cortex, while monkeys rested quietly, until we could isolate the spiking activity of individual neurons on the majority of

electrodes in each cortical area. We recorded local field potentials on up to 4 electrodes in each cortical area (LFPs, filtered 1-100 Hz) along with neural spiking activity concurrently. Once the spiking activity of a large number of individual neurons was evident in the neural recordings, we initiated the DPX task, and neural recording during task performance commenced. We recorded neural activity as monkeys performed blocks of 301 or 400 trials of the DPX task. We frequently recorded the activity of the same neural ensemble on multiple blocks, and on some days, advanced the electrodes to record additional neural ensembles.

We recorded neural ensemble activity in three experimental conditions relating to drug administration: Control (drug naive), Drug, and Saline. We first recorded neural activity in the Control condition for a number of days in each monkey, and then recorded neural activity in the Drug and Saline conditions on alternating days thereafter. Neural data in the Control condition was recorded before monkeys received their first injection of phencyclidine. Neural activity in the Drug condition was recorded after an intramuscular injection of phencyclidine (0.25-0.30 mg/kg i.m, Sigma Pharmaceuticals). Neural activity in the Saline condition was recorded after an intramuscular injection of an equivalent volume of sterile saline (after first injection of phencyclidine on a prior day).

In our prior report (3), we found that behavioral performance and spike timing dynamics differed between the Drug and Saline conditions when these conditions were interleaved over days. We also observed weaker persistent changes in neural dynamics evident in the Saline relative to the Control (drug naive) condition. This suggested that repeated exposure to PCP produced lasting effects on prefrontal network dynamics that persisted on Saline days. We therefore focus on neural activity recorded in the Control and Drug conditions, omitting Saline data to exclude potentially chronic effects of NMDAR blockade on neural interactions.

We restricted neural ensembles to the subset of neurons (typically about half of the neurons recorded) that exhibited significant modulations in firing rate in relation to task events. Task related neurons included those which modulated their firing rate in relation to the identity of the cue during the cue period, the cue or the probe during the probe period (some neurons active during the probe period encoded the prior cue), or the response during the response period. See (4) for additional details regarding statistical definitions of task-related activity. The present report is based on the spiking activity of 723 parietal and 1039 prefrontal task-related neurons. The Control condition consisted of 34 neuronal ensembles including 289 parietal neurons (average 11 per ensemble) and 468 prefrontal neurons (average 15 per ensemble). (For 13 of these control ensembles, monkeys received an injection of sterile saline before recording.) The Drug condition consisted of 34 parietal prefrontal ensembles including 434 parietal neurons (average 13 per

ensemble) and 571 prefrontal neurons (average 14 per ensemble). We generally recorded spiking and LFP data from both areas each day, but occasionally LFP recordings were restricted to only one brain area.

Causal discovery analysis

We performed all causal discovery analyses with a specific implementation of the Greedy Equivalence Search (GES)(5) algorithm, called the Fast Greedy Equivalence Search (FGES)(6). GES takes a single data table as input and returns a partially directed graph as output. The output graph is generated by a guided search within the space of all possible structures over the variables in the input data (here task, spiking, and LFP data), looking for the model that optimizes a model fit statistic. This procedure (described below) is correct in the large sample limit, performs well in simulations, and has been applied to other data sets in a similar domain.

Starting from a model with no edges, GES iteratively adds edges that most improve the model fit score until the score cannot be improved by any edge addition, and then removes edges until the score is not improved by any edge removal. This search procedure provably converges to the model with optimal model fit. The primary determinant of its output, then, is the model fit criteria chosen. We applied the widely used Bayesian Information Criterion (BIC) score (7), a penalized likelihood score. There are multiple equivalent formulations of the BIC score; we define it here as

$$\text{BIC} = -2 \ln(L) + \ln(n)\, k$$

where L is the likelihood of the data given the model, n is the sample size, and k is the number of free parameters in the model. Penalized likelihood scores attempt to balance the plausibility that a model could have generated the observed data, i.e. the likelihood L, against the complexity of that model, i.e. the number of free parameters k: unnecessarily complex models are typically capable of generating observed data as well or better than any model with fewer parameters. There are a number of competing penalized likelihood scores, but the BIC score enjoys the property of being optimized by the model that actually generated the data assuming the true model is among those tested.

To understand how this can happen, first note that the BIC score is the sum of two components: a log likelihood and a complexity penalty. A model's likelihood is defined as the probability of the data given the model; the log likelihood is simply the log of that value. Since all the models we consider are Bayesian networks, the likelihood is a product of values drawn from the Bayesian network's conditional probability tables:

$$P(X|M) = \Pi_{s \in S} \Pi_{v \in V} P_M(v = X_{s,v} | M_{pa}(v) = X_{s,M_{pa}(v)})$$

Where *X* is the data, *M* is the model, and *S* and *V* are the samples and variables. P_M is a multinomial probability distribution computed from the data and stored as conditional probability tables in model *M*, $X_{s,v}$ is the value of variable *v* in sample *s* (0 or 1 in our case), $M_{pa}(v)$ is the set of variables that are parents of *v* (e.g. send directed edges to *v*) in model *M*, and $X_{s,M_{pa}(v)}$ is the corresponding set of variable values in sample *s*. A key step is deriving the conditional probabilities P_M that all variables *v* in the model take given concurrent values in the parent variables ($M_{pa}(v) = X_{s,\, Mpa(v)}$). These conditional probabilities are computed directly from the data. For example, let *Y*, *A*, and *B* be variables in *V*, and let *A* and *B* be parents of *Y* in model *M*. Imagine that we observe in sample s that *Y*=1, *A*=1, and *B*=0. Then we compute

$$P_M(Y = 1 | A = 1, B = 0) = \frac{count_X(Y=1, A=1, B=0)}{count_X(A=1, B=0)}$$

where $count_X$ is simply a function that counts how many samples in *X* satisfy the condition in its argument. Once the conditional probabilities are derived, the likelihood of the data given the model is computed as the product of the conditional probabilities over all samples and variables.

The semantics of the edges in graphical representations of Bayesian networks is also contained in the way their likelihood is computed: variable A points at (is a parent of) variable B if A is used by the model in the calculation of the probability of B, otherwise A does not point at B. It also provides an intuition of how these graphs relate to, but should not be equated with, correlations. If A and B are correlated, the correlation will seem unlikely according to models without an active path between those nodes. Models with edges that connect A and B, either directly or indirectly, will have an advantage when computing their BIC scores, as they will assign a higher likelihood to such a correlation being observed.

The complexity penalty is a term with opposite sign to the log likelihood, and it grows in magnitude as the model becomes more complex. If variables A and B are uncorrelated, then models with edges connecting A and B will not have much better likelihood, but the added edges represent a complexity by introducing more free parameters to the model, increasing *k* and the penalty part of the score. Models with those unnecessary edges will have worse BIC scores than similar models without such edges. The model that optimizes the BIC score, then, is the model whose edges can explain all and only those correlations and conditional correlations that are

observed in the data: being able to explain anything more is punished by the complexity penalty, and not being able to explain enough is punished by the likelihood.

We used the FGES implementation from the open source software package Tetrad version 6.6.0 (http://www.phil.cmu.edu/tetrad/index.html), using the discrete Bayesian Information Criterion score as the model fit criterion (as all data were preprocessed to be binary and the sample size is large). We left the parameter for the discrete BIC score at its default value (penaltyDiscount = 2). We also set the maximum degree to -1 (unbounded), to avoid sacrificing accuracy for speed. FGES also allows its search of the space of all structures to be restricted by the user, intended for cases where the user knows that certain structures are impossible. Given the fixed nature of the task variables in the experimental design, we removed all models in which the task variables are caused by any other variables. The same settings were used for all applications of causal discovery analysis to the Control and Drug condition data.

Quantification of causal interactions

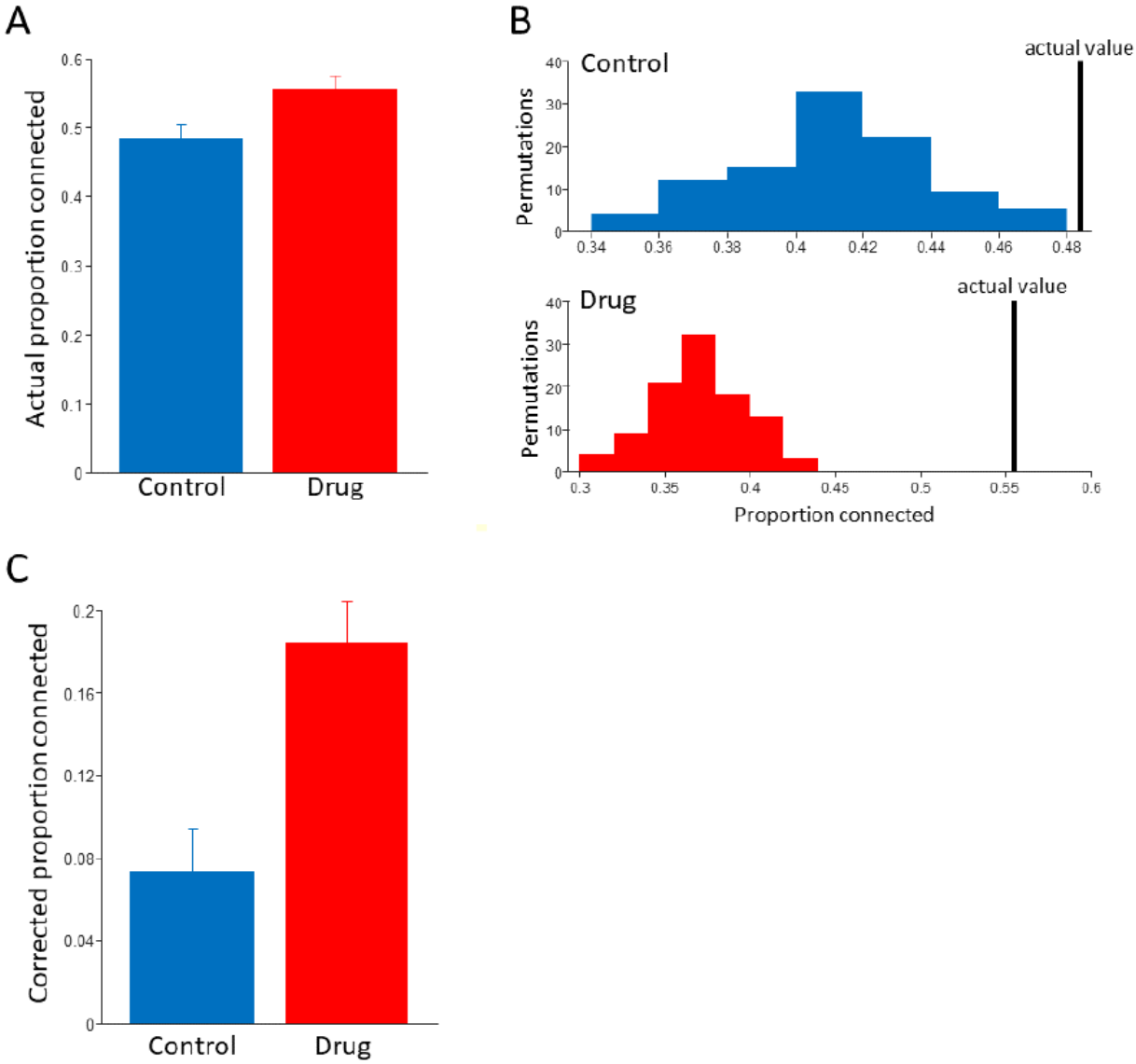
We applied the causal discovery analysis separately to 68 data sets (34 each in the Control and Drug conditions), providing us with 68 maps of connections between simultaneously recorded neurons and LFP data. To assess whether and how blocking NMDAR influenced the pattern of multiscale causal interactions, we compared the rates at which connections occurred between neuronal spiking and LFP data within and between cortical areas across all recorded neural ensembles, for both the Control and Drug conditions, relative to the possible number of connections available across these conditions. For example, a data set collected from prefrontal cortex with five LFP signals (one for each frequency band) and spiking data from three neurons would have a total of 40 possible within-area LFP-LFP connections (5x4 = 20 directional pairs x 2 directions), 12 possible spike-spike connections (3x2 = 6 possible directional connections x 2 directions), and 60 possible LFP-spike connections (2x5x3 = possible directional connections x 2 directions). Possible connections of signals across areas were tabulated similarly.

From such calculations we were able to obtain proportions of connections observed relative to those available. Since we were interested in interactions that reflected true functional (e.g. synaptically mediated) interactions between neural signals, and not neural events that fluctuate together during a trial as a function of visual stimuli or other external events, we normalized the proportions of connections we observed by subtracting the proportion of connections we expect to see between neural signals based chance or by entrainment of those signals to external events. To do this, we performed 100 permutation causal discovery analyses on each data set, with neural data shuffled randomly across trials, and for each permutation we computed the proportion of

connections detected relative to all possible connections as described above. This gave us permutation distributions for the proportion of connections detected when neural signals remained locked to external events but were not simultaneously recorded. We corrected the proportions of connections observed in the simultaneous data by subtracting the means of the corresponding permutation distributions. (See Supplemental Figure 1 for an example of this correction procedure.) To determine whether the corrected proportions were significantly different between Control and Drug conditions we compared the actual difference in corrected proportions ("corrected" by subtracting the mean of the permutation distribution) to a permutation distribution of 100 differences between Control and Drug conditions obtained after shuffling trials. The p value of the actual difference was calculated as the proportion of permutation differences exceeding the actual difference.

We also compared the results of our causal modeling analysis across different trial types (correct AX, correct BX, and error BX trials). In this analysis, we restricted our data to only the Drug condition, since error rates were too low in the Control condition to provide enough error trials. We included any data set that had at least 10 BX error trials, and randomly downsampled the correct trials (both AX and BX) in that data set to match the number of BX error trials. This analysis included 31 data sets, ranging from 13 to 51 trials (mean of 23 trials).

Supplemental Figure



Supplemental Figure S1. Proportion correction procedure for parietal to prefrontal macroscale LFP-LFP connections. **A.** Actual proportions of connections observed in the two drug conditions, relative to the total possible set of connections across data sets in each condition. In the Control condition, we observed 254 connections out of a total of 525 possible connections (48%). We observed 347 of 625 possible connections in the Drug condition (56%). Error bars indicate the standard error of proportion. These proportions are significantly different from each other, using a z test of proportions ($p < 0.05$). **B.** Permutation distributions of proportions of connections. We ran the causal discovery modeling analysis 100 times per data set, shuffling data points across trials. We then calculated the proportion of parietal to prefrontal LFP connections at each of these 100 iterations, relative to the total possible. We found that in this case, the actual values we observed in the data exceed all values obtained in the shuffled data, in both the Control (top) and Drug (bottom) conditions. (We found this to be the case typically–nearly all actual values exceeded the 95th percentile of the permutation distributions.) **C.** Corrected proportions. We calculated the corrected proportion of connections for each drug condition by subtracting the mean of the permutation distribution from the actual proportion observed. Error bars in C indicate the standard deviation of the permutation distributions shown in B.

Supplemental References

1. Oostenveld R, Fries P, Maris E, Schoffelen JM (2011): FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*. 2011:156869.:10.1155/2011/156869. Epub 152010 Dec 156823.

2. Blackman RK, Macdonald AW, 3rd, Chafee MV (2013): Effects of ketamine on context-processing performance in monkeys: a new animal model of cognitive deficits in schizophrenia. *Neuropsychopharmacology*. 38:2090-2100. doi: 2010.1038/npp.2013.2118. Epub 2013 May 2010.

3. Zick JL, Blackman RK, Crowe DA, Amirikian B, DeNicola AL, Netoff TI, et al. (2018): Blocking NMDAR Disrupts Spike Timing and Decouples Monkey Prefrontal Circuits: Implications for Activity-Dependent Disconnection in Schizophrenia. *Neuron*. 98:1243-1255.e1245. doi: 1210.1016/j.neuron.2018.1205.1010. Epub 2018 May 1231.

4. Blackman RK, Crowe DA, DeNicola AL, Sakellaridi S, MacDonald AW, 3rd, Chafee MV (2016): Monkey Prefrontal Neurons Reflect Logical Operations for Cognitive Control in a Variant of the AX Continuous Performance Task (AX-CPT). *J Neurosci*. 36:4067-4079. doi: 4010.1523/JNEUROSCI.3578-4015.2016.

5. Rosch RE, Auksztulewicz R, Leung PD, Friston KJ, Baldeweg T (2019): Selective Prefrontal Disinhibition in a Roving Auditory Oddball Paradigm Under N-Methyl-D-Aspartate Receptor Blockade. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 4:140-150. doi: 110.1016/j.bpsc.2018.1007.1003. Epub 2018 Aug 1013.

6. Ramsey J, Glymour M, Sanchez-Romero R, Glymour C (2017): A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *Int J Data Sci Anal*. 3:121-129. doi: 110.1007/s41060-41016-40032-z. Epub 42016 Dec 41061.

7. Schwarz G (1978): Estimating the dimension of a model. *Ann Stat*. 6:461-464.