SUPPLEMENT

Schizophrenia is Associated with a Pattern of Spatial Working Memory Deficits Consistent with Cortical Disinhibition

Martina Starc[1,2]*, John D. Murray[1,3]*, Nicole Santamauro[1], Aleksandar Savic[1,4], Caroline Diehl[1,2], Youngsun T. Cho[1], Vinod Srihari[1], Peter T. Morgan[1], John H. Krystal[1,8], Xiao-Jing Wang[3] , Grega Repovs[2]*, Alan Anticevic[1,5,6,7,9]*

1. Department of Psychiatry, Yale University School of Medicine, 300 George Street, New Haven, CT 06511, USA
2. Department of Psychology, University of Ljubljana, Ljubljana, Slovenia
3. Center for Neural Science, New York University, New York, NY 06510, USA
4. University Psychiatric Hospital Vrapce, University of Zagreb, Zagreb 10000, Croatia
5. Abraham Ribicoff Research Facilities, Connecticut Mental Health Center, New Haven, CT 06519, USA
6. Interdepartmental Neuroscience Program, Yale University, New Haven, CT 06520, USA
7. NIAAA Center for the Translational Neuroscience of Alcoholism, New Haven, CT 06519, USA
8. Department of Neurobiology, Yale University, New Haven, CT
9. Department of Psychology, Yale University, 2 Hillhouse Avenue, CT 06520, USA

* designates equal contributions

Short Title: Disinhibition and Working Memory in Schizophrenia

Corresponding Author:

Alan Anticevic Ph.D., Yale University, Department of Psychiatry
34 Park St., New Haven, CT 06519, Office (203) 974-7763
alan.anticevic@yale.edu

Total Words: ~4500
Abstract: 236
Tables / Figures: 1 / 5

FINANCIAL CONFLICTS OF INTEREST: J.H.K. consults for several pharmaceutical and biotechnology companies with compensation less than $10,000 per year. He also has stock options in two companies, each valued less than $2,000 and three patents for pharmacotherapies for psychiatric disorders. None of these financial interests are directly related to this paper. All other authors declare that they have no conflict of interest.

Complete Working Memory (WM) Experimental Paradigm Details. Of note, the WM paradigm was deigned to explicitly mimic the 'ring' architecture of the computational model (Compte et al., 2000), in order to directly test model-derived predictions. Subjects were instructed to remember the cue position and maintain its location over variable delay of 0s, 5s, 10s, 15s, or 20s (with 60, 50, 40, 30 and 20 trials respectively). Here different number of trials per delay duration were implemented in order to ensure that subjects were equally likely to get tested at every particular point in time, thus encouraging continued maintenance and response expectations throughout the delay period. Put differently, if there were equal numbers of long and brief trials, then subjects could implicitly learn that after a few seconds they are guaranteed to get a longer trial as they occur at equally high frequency. The final number of trials at different delays was a compromise between the objective to achieve equal probability that a response will be needed at each time point, and obtaining a sufficient number of trials at each duration for statistical analyses (i.e. power analyses considerations). If there were equal number of trials (40) of each duration, there would be only 20% (40/200) probability that a participant will have to respond immediately vs. later on, as the trial duration gets longer. At 5 seconds that probability would increase to 25% (40/160), 33% (40/120) at 10 s delay, 50% (40/80) at 15 s and 100% (40/40) at delay 20. This might induce the participant to reduce their attention allocation at the start and increase attention throughout the progression of the trial. To ensure equal probability of response at each time point, the number of trials of a specific delay would need to diminish exponentially, with 100 trials of delay 0, and 50, 25, 13, and 12 trials at delays 5, 10, 15 and 20s respectively. This would significantly reduce the number of useful trials at long durations and consequently attenuate the statistical power and broaden the confidence intervals. The compromise distribution we used 'flattened' the probability of responses across trial progression—30, 36, 33, 60, 100% probability of response at delays 0, 5, 10, 15 and 20 s respectively—while still ensuring enough trials of the longest duration.

After the delay period a grey circle of the same size as the cue (125px diameter, 2.8s) appeared in the middle of the screen (see **Figure 1**). Subjects were instructed to move the probe to the remembered position with a high-sensitivity 16-bit joystick and hold it at the remembered location until the probe disappeared. The final X/Y position of the gray circle was recorded as a continuous response. The delay task tested whether the SCZ group would exhibit greater WM response variability as a function of delay duration (i.e. WM drift) relative to the HCS group.

The distractor WM task was identical to the delay WM task in its overall structure but contained only one delay duration (10s), during which an additional red circle (the distractor) was presented (125px diameter, 1.4s). The distractor was presented in the middle of the delay duration (after 4.3s). Subjects were instructed to keep maintaining the position of the cue and keep their eyes fixed on the middle of the screen throughout the distractor period (see **Limitations** for considerations surrounding eye tracking). In this experiment two distractor distances were chosen, at 20° (proximal) and 50° (distal) from the cue location. Twenty cue positions were used and each was presented twice for a total of 40 trials per distractor distance. The 10s delay condition from the delay WM task served as a no-distractor control condition. Again, the distractor task tested whether the SCZ group would exhibit a differential response bias in the direction of the distractor across distractor distances relative to the HCS group, as predicted by the model.

Subjects also completed a control 'motor' response task (20 trials) where both the cue and the probe appeared simultaneously and subjects were instructed to cover the cue with the probe circle, which necessitated a motor response but no WM maintenance or recall. This control task was included to verify that potential differences between groups are not exclusively driven by significantly lower motor skill in the SCZ group.

Experiments were run on an LCD screen (1280×1024px resolution). As noted, a fixation point was present on the middle of the screen throughout the trial and subjects were encouraged to maintain fixation at all times. 20 different cue positions were used in a balanced design (starting at 0° angle and evenly spaced around the invisible ring in 18° steps). In the distractor task, 20 cue positions were used but were shifted 10° counter-clockwise compared to the delay task to exclude angles representing cardinal axes. Collectively, the WM tasks contained 220 trials per subject.

Percentage of Excluded Trials Due to Lack of Adequate Response. To maintain quality control, we discarded trials where subjects failed to initiate a response. These trials were defined based on two criteria: i) trials where the probe was moved <1/3 the way to the rim of the accurate location; ii) trials that were displaced >90° degrees in any angle from the cue position. These trials were treated as outliers and reflected lack of adequate responses (2% of all trials). These trials likely reflect i) a drop in attention that causes a failure to initiate a motor response in time resulting in responses close to the starting point, or ii) a complete failure of WM resulting in a very large angular displacement (missing the correct half of the screen). To ensure our main analyses were not influenced by exclusion of these outliers across groups we tested for differences in the percent of excluded trials between groups and trial conditions with the same mixed-measures ANOVAs used in the main analyses (see **Methods** for details).

Results revealed a main effect of *Diagnosis* [F(1,53)=8.04, p=.006], and a main effect of *Delay Duration* [F(5, 265)=3.915, GGe=0.605, p=.010], but no significant *Delay Duration* x *Diagnosis* interaction [F(5,265)=1.492, GGe=0.605, p=.218 (**Figure S1**). The effect of *Delay Duration*, however, was driven by the motor condition. Removing it from the analysis eliminated the *Delay Duration* effect [F(4,212)=1.049, GGe=0.592, p=.362]. The difference between groups remained significant [F(1,53)=7.137, p=.010]. These results indicate that SCZ patients exhibit more attentional or WM lapses in general. However, no consistent increase of these errors with increasing WM delay duration, ruling out the possibility that excluding these outlier trials has biased the main delay effects presented in the main text. Finally, when repeating this analysis for the WM distraction task we found no significant effects for outlier trials: no effect of *Diagnosis* [F(1,50)=0.137, p=.713], no effect of *Distractor Location* [F(2,100)=0.01, GGe=0.593, p=.948] and no *Diagnosis* x *Distractor Location* interaction [F(2,100)=1.056, GGe=0.593, p=.321] (**Figure S2**).

Percentage of All Excluded Trials. Additionally, z-values of angular displacement were calculated for each subject in each condition and all results with absolute z-values greater than 3 (an additional 1% of all trials) were removed to ensure outliers are not driving reported effects. Testing for differences in the percent of all excluded trials yielded no significant effects both for the delay WM task: no effect of *Diagnosis* [F(1, 53) = 1.353, p = .250], no effect of *Delay Duration* [F(5,265)=0.995, GGe=0.648, p=.401] and no *Diagnosis* x *Delay Duration* interaction [F(5, 265)=2.082, GGe=0.648, p=.099] (**Figure S3**). In addition, there were no differences in the proportion of excluded outlier trials for the distraction WM task: no effect of *Diagnosis* [F(1,50)=0.002, p=.961], no effect of *Distractor Location* [F(2,100)=0.433, GGe=0.659, p=.567], and no *Diagnosis* x *Distractor Location* interaction [F(2,100)=0.217, GGe=0.659, p=.710] (**Figure S4**). Collectively, these control analyses strongly support the conclusion that exclusion of outlier trials did not significantly influence our main analyses.

Testing Effects of Delay Duration on Spatial WM Performance in a Subsample of Matched Trials. In the delay task, different number of trials were included for each delay duration (0s: 60, 5s: 50, 10s: 40, 15s: 30 and 20s: 20 trials). As noted, this was done in order to ensure that subjects were equally likely to get tested at

every particular point in time, thus encouraging continued maintenance and response expectations throughout the delay period. It is vital to control for 'expectation' of a longer trial duration to circumvent possible differential rehearsal strategies. Put differently, if the number of trials was equal across all delay periods then the probability for a longer period trial would increase with each subsequent second passing. The way to control for this involves explicitly re-balancing the number of trials at longer delays to comprise a smaller number of overall trials, thus reducing an implicit expectation bias ever forming. While this control is balanced across groups, this psychometric control may effectively introduce a differential level of 'sensitivity' across delay periods due to far greater trial numbers. We sought to explicitly rule out this possibility by randomly sampling an equal number of trials across all delay durations (one trial for each of the 20 used locations), and then repeating all main analyses.

As in the full analyses (see **Figure 3**), here we first examined whether SCZ is associated with greater variability of response accuracy at the probe phase, especially as a function of increased delay duration. Response patterns in **Figure S5A** indicate that the spread of WM responses increases with delay duration across groups. However, SCZ patients exhibited wider spread at all levels of delay. Furthermore, the spread increased more for the SCZ group with longer delays, as with the overall analysis. We tested this effect formally by calculating the SD of angular displacement (**Figure S5B**), which revealed a significant main effect of *Delay Duration* [$F(5,265)=133.829$, GGe=0.575, $p<.001$], reflecting increased response variability as a function of longer WM delays. We again observed a significant main effect of *Diagnosis* [$F(1,53)=11.56$, $p=.001$], indicating greater variability overall in the SCZ group (Cohen's d 0.61–1.10, greatest at 20s delay, **Figure S5D**). Critically, as before, results revealed a significant *Delay Duration x Diagnosis* interaction [$F(5,265)=5.297$, GGe=0.575, $p=.002$], indicating a steeper slope of WM drift as a function of delay duration in SCZ, as reported for the main analyses (see **Figure 3**). In turn, we again calculated that for each subsequent second of the delay period the angular displacement SD increased on average for 0.14 degrees in HCS group and 0.26 in SCZ group (average subject slope, [$t(51.19)=3.537$, $p<.001$, one-tailed], Cohen's d = 0.95, **Figure S5E**). Importantly, this elevated change in WM variability was again not associated with a particular directional bias. Put differently, variability of responses increased across both groups (**Figure S5B**) but the average distribution remained centered on the WM cue (**Figure S5C**), indicating a 'random' spatial drift. This was also confirmed statistically as we observed no significant angular bias effects: no effect of *Diagnosis* [$F(1,53)=2.139$, $p=.149$], no effect of *Delay Duration* [$F(5,265)=0.227$, GGe=0.557, $p=.864$], and no *Delay Duration x Diagnosis* interaction [$F(5,265)=1.032$, GGe=0.557, $p=.376$]. Collectively, these effects strongly support the conclusion that greater number of trials across the WM delay periods did not impact our core results.

Effects of Education and Intelligence on Spatial WM Performance in the Delay Task. Our samples predictably differed in educational attainment, verbal and non-verbal intelligence given that SCZ illness course significantly impacts these variables. When considered in isolation, all variables were significant predictors of performance in the delay task: education [$F(1,53)=8.363$, $p=.006$], verbal intelligence [$F(1,50)=12.554$, $p< .001$; non-verbal intelligence [$F(1,51)=19.74$, $p<.001$]. However, when these possible confounding variables were explicitly modeled as a covariate, the main effect of education was no longer significant [$F(1,52)=1.302$, $p=.259$], but it attenuated somewhat the main effect of *Diagnosis* [$F(1,52)=3.935$, $p=.05$]. Nonetheless, the main effects of *Delay Duration* [$F(5,265)=150.665$, $p<.001$], and the *Diagnosis* x *Delay Duration* interaction [$F(5,265)=6.155$, $p<.001$], remained highly significant even when explicitly co-varying for education levels. Verbal intelligence remained significant when included as a covariate [$F(1,49)=4.743$, $p=.034$], and also

attenuated the main effect of *Diagnosis* [F(1,49)=3.327, p=.074]. The main effect of *Delay Duration* [F(5,250)=137.87, p<.001], and the *Diagnosis* x *Delay Duration* interaction [F(5,250)=5.708, p<.001], remained highly significant when explicitly co-varying for verbal intelligence levels. Similarly for non-verbal intelligence, the covariate showed a significant main effect [F(1,50)=9.96, p=.003], and attenuated the main effect of *Diagnosis* [F(1,50)=1.691, p=.199], while the main effect of *Delay Duration* [F(5, 255) = 140.311, p < .001], and the *Diagnosis* x *Delay Duration* interaction [F(5,255)=5.562, p<.001], remained highly significant. Importantly, after regressing out the effects of covariates, and calculating the slopes of WM drift with time on residual values, differences between groups in slope steepness remained significant even when controlling for all three covariates: education [t (51.58)=3.73, p<.001, one-tailed], verbal intelligence [t (44.85)=3.46, p<.001, one-tailed], and non-verbal intelligence [t (47.3)=3.495, p<.001, one-tailed]. Collectively, these control analyses do not support the possibility that the *Diagnosis* x *Delay Duration* interaction, a key predicted effect, is explained by secondary variables likely affected by the illness course.

Effects of Education and Intelligence on Spatial WM Performance in the Distraction Task. As with the delay analyses above, we sought to rule out the possibility that educational attainment or intelligence per se were driving the core reported effects. Here we found that educational attainment [F(1,50)=4.145, p=.047] and non-verbal intelligence [F(1,50)=15.661, p<.001] both significantly predicted angular displacement in the distraction task. Verbal intelligence did not reach significance [F(1,49)=3.651, p=.062]. Including education as a covariate did not alter the pattern of results. While its effect was reduced when included as a covariate [F(1,49)=1.386, p=.245], there was no main effect of *Diagnosis* [F(1,49)=0.633, p=.430], but the main effect of *Distractor Location* [F(2,100)=24.705, p<.001], and *Diagnosis* x *Distractor Location* interaction [F(2,100)=4.032, p=.021] both remained highly significant. A similar pattern was found for verbal intelligence. There was no significant effects of the covariate [F(1,48)=1.362, p=.249], or *Diagnosis* [F(1,48)=0.757, p=.389], but we still observed a significant main effect of *Distractor Location* [F(2,98)=24.046, p<.001], and a significant *Diagnosis* x *Distractor Location* interaction [F(2, 98)=3.995, p=.021]. Only non-verbal intelligence remained a significant predictor of angular displacement after inclusion as a covariate [F(1,49)=11.325, p=.001]. However, the main results remained unaltered: main effect of *Diagnosis* was attenuated as with other covariates [F(1,49)=0.001, p=.974], whereas *Distractor Location* [F(2,100)=24.705, p<.001] and *Diagnosis* x *Distractor Location* [F(2,100)=4.032, p=.021] both remained significant when including non-variable intelligence as a covariate. As with the delay effects, collectively, these control analyses do not support the possibility that the *Diagnosis* x *Distractor Location* interaction is explained by these secondary variables.

Effects of Medication on Spatial WM Performance. As reported in the main text, the majority of patients were receiving antipsychotic pharmacotherapy, which we converted to chlorpromazine (CPZ) equivalents and correlated with performance on the two WM tasks. Including medication as a covariate in the analysis of the delay task showed no significant effects on variability of angular displacement [F(1,21)=0.145, p=.707]. In addition, the main effect of *Delay Duration* remained significant in the SCZ group [F(5,110)=97.105, p<.001]. Similarly, when examining the distraction WM task, medication as a covariate had no effect on angular displacement [F(1,19)=0.205, p=.655]. Also, the main effect of distance in the SCZ group remained significant [F(2,40)=10.523, p<.001]. Collectively, these control analyses cannot rule out the possibility that results would perhaps differ in fully unmedicated patients but they do rule out the possibility that the aggregate level of current medication dose is related to core reported effects in SCZ patients.

Testing for Effects of Distractor Position of Spatial WM Performance in a Subsample of Subjects. As described in the main text, we tested whether the response distribution in the distal distractor condition for SCZ subjects is bi-modal, with a fraction of responses centered on the target and a fraction centered on the distractor. Testing at the group or individual level, via the Hartingans' dip test, did not reveal a bi-modal distribution. However, due to the relatively small number of responses in each condition (N = 40), the test might not have significant power to detect bimodality of distribution in each individual subject. There is a possibility that on some trials, subjects do forget the target and instead start maintaining the position of the distractor and respond with a large error or a response that is closer to the distractor than the target. However, it is challenging to fully determine which responses fall into this category, because the observed spread of responses is fairly close to the distracter position even in the distal distractor condition. To fully rule out the possibility that such a response pattern is not driving the differences in means observed between groups, we verified whether some subjects in the distal distraction condition show an increased percentage of responses that are closer to the distractor than to the target (angular displacement > 25°). This pattern would be consistent with potentially a complete loss of the target memory trace. Of the 25 patients with distraction data, two exhibited a marked effect with almost 50% of responses closer to the distractor than to the target. One patient had 10% of such responses, and three about 3% of such responses. Additionally, two control subjects also exhibited about 3% of such responses. We implemented a stringent criterion, which excluded all the outlier subjects while retaining those subjects that never crossed the middle line between target and distractor in the distal distractor condition. This resulted in a sub-sample of 25 HCS and 19 SCZ subjects.

Next, we assessed if the general pattern of increased distractibility in SCZ patients in the distal distractor condition still holds in this subsample of subjects that did not exhibit a tendency to respond with the position of the distractor in any of the trials. The increased spread of responses in the SCZ group was still present (*Diagnosis* main effect [$F(1, 42) = 6.947$, $p = .012$]) (**Figure S6A-B**). Results also revealed a main effect of *Distractor Location* [$F(2, 84) = 9.258$, $GGe = 0.8$, $p < .001$], but no *Diagnosis* x *Distractor Location* interaction [$F(2, 84) = 0.741$, $GGe = 0.8$, $p = .452$]. In the proximal distractor condition, the spread of responses was again non-specific. However, there was a net bias toward the distractor for the distal condition. Therefore, in addition to response variability, distractors affected the average angular displacement of response location (**Figure S6C**). Proximal distractors were associated with an angular shift away from the distractor in both groups equally while distal distractors were associated with a shift toward the distractor in SCZ patients only [$t(32.72) = -2.615$, $p = .040$]. The pattern of ANOVA effects did change somewhat in this subgroup with no main effect of *Diagnosis* [$F(1, 42) = 1.197$, $p = .280$] and a significant main effect of *Distractor Location* [$F(2, 84) = 69.649$, $p < .001$]. The *Diagnosis* x *Distractor Location* was significant at the trend level [$F(2, 84) = 3.037$, $p = .05$]. This likely reflects the fact that numerically, the 'attractive' distractor effect in the distal condition (where there are differences between groups) is smaller than the 'repulsive' distractor effect in the near condition where both groups are affected equally (whereas the opposite is the case in the full sample). That said, the difference between groups in the distal distractor condition remains significant and the effect size is even larger in this subsample of subjects despite smaller sample size (Cohen's $d = 0.80$, compared to 0.59 in the full sample).

SUPPLEMENTAL FIGURE CAPTIONS

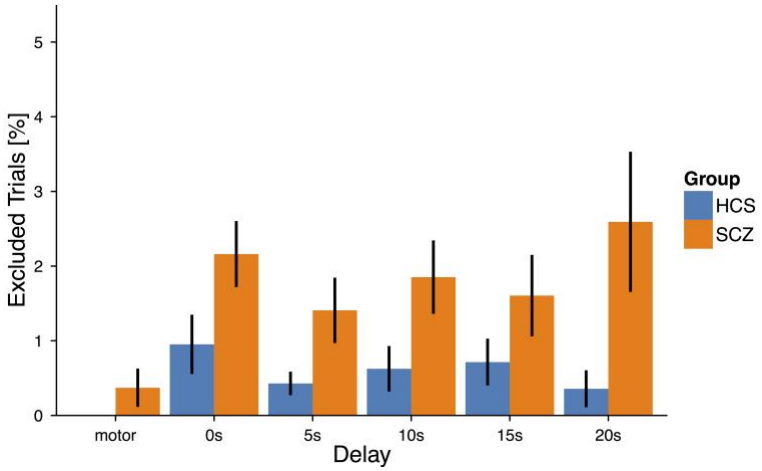

Figure S1. Percentage of Excluded Trials Based on Lack of Adequate Responses in the WM Delay Task. SCZ patients produce a greater percentage of inadequate responses at all delay durations, but there is no obvious pattern of increase with delay. Error bars show +/- 1 standard error of the mean.

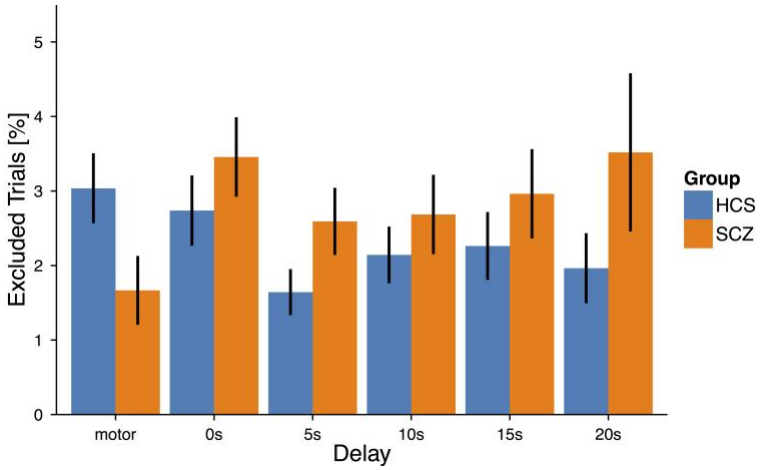

Figure S2. Percentage of All Excluded Trials in the WM Delay Task. Approximately equal percentages of trials are excluded in all conditions in both groups. Error bars show +/- 1 standard error of the mean.

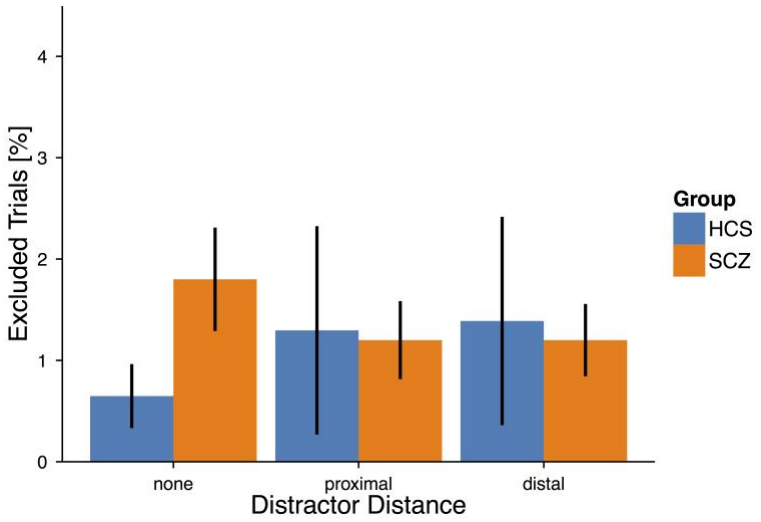

Figure S3. Percentage of Excluded Trials Based on Lack of Adequate Responses in the WM Distraction Task. Approximately equal percentages of trials are excluded in all conditions in both groups. Error bars show +/- 1 standard error of the mean.

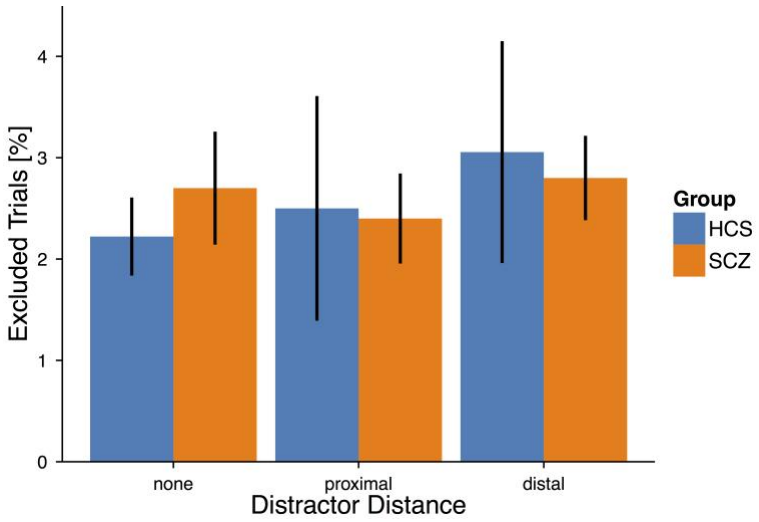

Figure S4. Percentage of All Excluded Trials in the WM Distractor Task. Approximately equal percentages of trials are excluded in all conditions in both groups. Error bars show +/- 1 standard error of the mean.

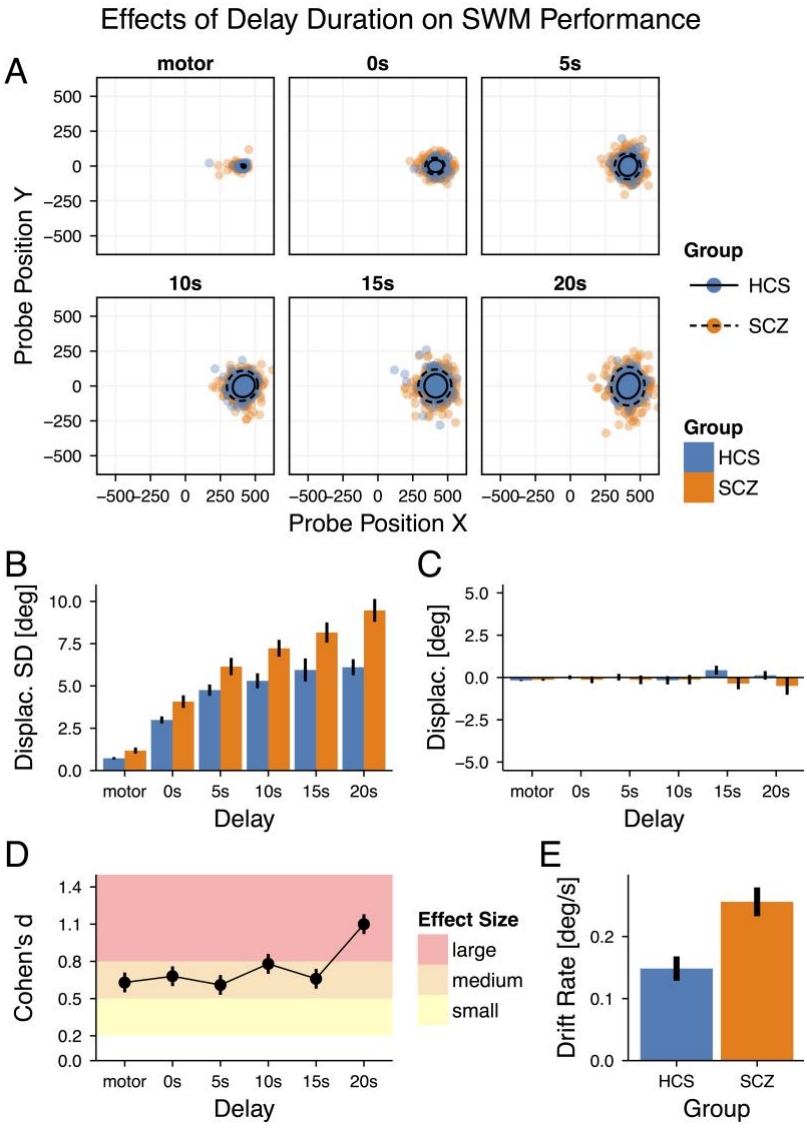

Figure S5. Effects of Delay Duration on Working Memory Drift with a Subsample of Trials. Subsampling trials (20 per delay duration) does not change results (compare to **Figure 3** in main text). **A.** Each panel indicates positions of responses on the screen (1280x1024px) at different delays. Results are rotated as if every cue was presented at 0^0 angle (x=415px, y=0px) to facilitate visual inspection. Gray dots indicate the positions of targets on the screen before rotation. 95% confidence ellipses are also shown around the mean response pattern for each group. Responses spread out with increasing delay but remain centered on the cue position. **B.** Average standard deviation of response errors measured as angular displacement from the cue location. Average variability is increased for the SCZ group compared to HCS [F(1,53)=11.56, p=.001] and this difference gets larger with increasing *Delay Duration* [F(5,265)=5.297, GGe=0.575, p=.002], indicating lower WM precision. **C.** As expected, average angular displacement of the responses is unaffected by either *Delay Duration* [F(5,265)=0.227, GGe=0.557, p=.864] or *Diagnosis* [F(1,53)=2.139, p=.149], indicating no directional bias in the response pattern, as predicted by the model. **D.** Effect sizes for the differences between groups shown in B are in the medium to large range and remain relatively stable as delay increases. The only notable increase in effect size happens at the longest delay. **E.** Average drift rate across the five delays shown in B is increased for the SCZ group compared to HCS [t (51.19)=3.537, p<.001, one-tailed]. Error bars show +/- 1 standard error of the mean.

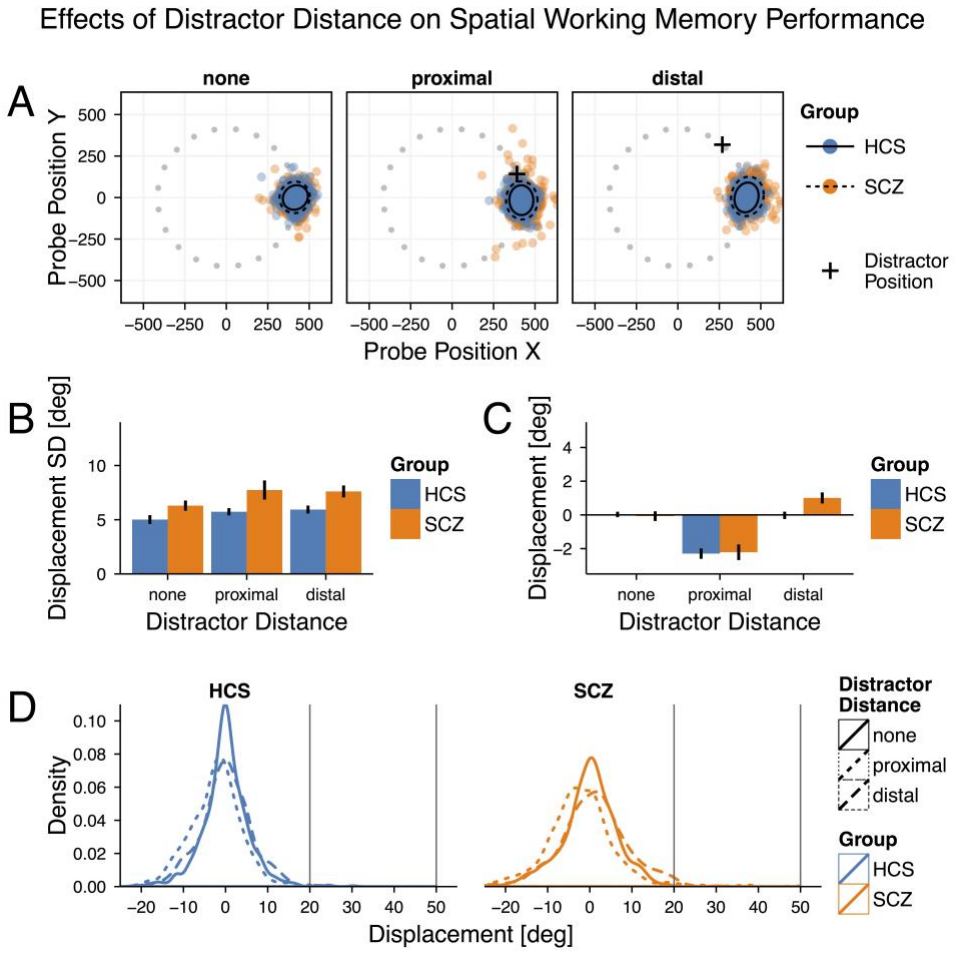

Figure S6. Effects of Distractor Position on Working Memory Performance. A. Each panel indicates positions of responses on the screen rotated to angle 0 with 95% confidence ellipses and gray dots marking target positions before rotation, as in Figure 3. Responses for the SCZ group show spreading in the presence of distractors, displaced away from cues toward distractors (crosses), in the distal distraction condition. **B.** Average standard deviation of response errors. Variability is increased in the SCZ group compared to HCS and in both groups in conditions with distractors. **C.** Distractors cause angular displacement of responses. While proximal distractors bias both groups to move away from the distractor, distal distractors only affect the SCZ group by biasing their responses toward the distractor location. Error bars show +/- 1 standard error of the mean. **D.** Density plots for HCS (left) and SCZ (right) illustrating the shifts in responses. Critically, these density plots are not consistent with bi-modal patterns of responses, but rather indicate subtle general right distribution shift for SCZ following distal distractors.