Supplementary Methods

Inclusion/Exclusion Criteria for Participants

Participants were primary English speakers, ages 18-55, devoid of history of head trauma and neurological disorders, and had normal or corrected visual acuity (20/40 or better). Participants with contraindication for MRI (artificial implants and/or claustrophobia) were excluded from the study. Patients met DSM-IV criteria for schizophrenia, as confirmed by the Lieber Schizophrenia Research Clinic (LSRC) before recruitment and participation. Patients were on disease-appropriate medication used within appropriate guidelines, and were assessed with the Positive and Negative Symptoms Scale (PANSS; (Kay, Fiszbein, & Opler, 1987)). Patients scoring above 120 on the PANSS, or those who were acutely psychotic were excluded. HC were recruited through the LSRC and through IRB-approved flyers and internet advertisements. The Structural Clinical Interview for the DSM-IV (SCID-NP) was used to exclude past or present Axis I or II disorders, significant substance use disorders in the past 6 months, and in HC significant psychiatric family history (First, Spitzer, Gibbon, & Williams, 1997).

TASIT Presentation and Eye-tracking

We showed participants the 32 video clips (23.0° x 19.1°) in TASIT Part 3 while they were seated 72cm from a 27" monitor (BenQ XL2720Z, BenQ USA, Costa Mesa CA), with their heads resting comfortably in a chin/head rest. TASIT Part 3 consists of an A and B section, each section containing 8 clips in which the main character actor is instructed to be sarcastic (using exaggerated affect) and 8 in which they are instructed to lie (using flat affect). The B section contains the same scenarios with the same dialogue as A, but with the sarcasm/lie instructions flipped. After each video clip, a research assistant verbally asked participants 4 questions about what the main character was thinking, feeling, doing, and saying, and recorded their responses with no time limit restriction. Only TASIT 3-A data were considered in this study due to the possibility of memory effects on TASIT 3-B performance. An eye-tracker (EyeLink 1000plus, SR Research, Mississauga, Ontario, Canada) recorded eye-movements at 1000Hz.

Statistical Analyses

All behavioral data were processed and analyzed by signal processing and statistical routines available in MATLAB (Mathworks, Natick MA). Magnitude values in the text are mean(standard deviation).

Visual Scanning Analyses

For all visual scanning analyses, each participant's eye-trace was first down-sampled to the video frame rate. For each video clip, participants with <50% of video frames with valid eye-tracking data were excluded to minimize instability in this measure.

Prediction of TASIT Performance

All regression analyses were repeated with antipsychotic dose in CPZ equivalents. To correct for differences in age between the two groups, the youngest 5 HC and oldest 5 SzP were removed from the sample, and the main analyses was repeated.

Visual Feature Classification

For each video, faces were detected using a combination of automated face classification routines (Zhu & Ramanan, 2012) that demarcated a box around each face in each frame of the video (see **Figure 3A**), along with linear interpolation to fill in face locations dropped by the automated algorithms (with some manual editing). Low level visual features (luminance, contrast, motion) were classified by first dividing the video into approximately 1°x1° cells (25x21 cells/frame) and then using the algorithms described in Russ *et al.* (Russ & Leopold, 2015) for each cell to create a continuous regressor for that cell for luminance, contrast, and motion (see **Figure 3B**).

The visual features in each cell were then averaged over a lagging 4 video frame (133ms) window to simulate the aggregation of visual information in visual cortex neurons. To simulate the relative (versus absolute) encoding of visual feature strength in visual areas, the visual feature intensities were normalized by the maximum visual feature intensity for that video frame. We then filtered these processed data through a neurophysiologically based visual field model that simulated the central versus peripheral visual field processing bias in visual cortex. To quantify visual features within faces, we masked non-normalized versions of the low-level visual feature maps by the face boundaries to create face-masked versions of the visual feature maps; these face-masked maps were then averaged of the 4 frame interval and filtered through the visual field model as above.

Visual Field Model

To quantify what visual features each participant was looking at, we first modeled each participant's visual field as the V4 cortical magnification factor modeled by Sereno and colleagues centered at the gaze position, with the values at visual angle <0.5° filled in with the values at 0.5° {Sereno:1995p997} (see **Figure 3A**). V4 was chosen as it is typically the cortical area studied in free-view visual search (Gee, Ipata, & Goldberg, 2010; Mazer & Gallant, 2003; Squire, Noudoost, Schafer, & Moore, 2013) and it has been shown to be under top-down attentional control (Armstrong, Fitzgerald, & Moore, 2006).

Quantifying Contents of Visual Field

For each frame, the visual field map was multiplied by the visual feature map (resampled to the original video resolution). Then, the surface integral of this multiplied map was used to quantify the combination of 1) the overlap of the visual field and the visual feature and 2) the intensity of that visual feature. In other words, the presence of the visual feature at the center (fovea) of the visual field would result in a high score (see **Figure 3C**), but also the strong presence of a visual feature near the visual field fovea would also receive a high score. For each visual feature, frames in which there were no faces present or that corresponded to the beginning of the video clip (a black screen with just the clip number displayed in yellow) were excluded.

Analysis of Saccades

Saccades and blinks were automatically detected by the eye-tracking software (SR Research, Mississauga, Ontario, Canada). From the automatically detected saccades, those with durations more than 300ms and/or amplitudes <0.25° were then removed.

Analyzing Saccades to Visual Features within Faces

To analyze which low-level visual features drove saccades to a face, the following steps were performed. 1) The automated algorithm in the eye-tracker, supplemented with additional artifact removal processing (see above), was used to detect and measure the amplitude of saccades. 2) Saccades were separated by whether they were to faces or not. 3) We measured the visual feature strength within the targeted face in the 4 video frames (133ms) before the beginning of the saccade. 4) We plotted the number (density) of saccades made to the faces as a function of both saccade amplitude and the visual feature strength. Due to the sparse number of saccades made to each 0.25° bin, saccades density plots were combined for each group. 5) The resulting 2D histograms were smoothed and subtracted between groups. 6) The resulting difference maps were thresholded at the 90th percentile. 6) We then searched for clusters of high saccade

count density at saccade amplitudes >5°
(https://www.mathworks.com/matlabcentral/fileexchange/57669-find-clusters).

Supplementary Results

Control Analyses for TASIT Performance Prediction

Antipsychotic medication dose in chlorpromazine equivalents also did not contribute significantly (p=0.9) to SzP TASIT sarcasm performance. The results described above remained significant even after controlling for group age differences by removing the five youngest HC and 5 oldest SzP and equating group ages (HC versus SzP, 37.8yo versus 38.5yo, t_{54}=-0.26, p=0.8, **Supplemental Tables 1** and **2**).

The MATRICS Consensus Cognitive Battery (MCCB) was also collected in SzP (excluding the MSCEIT subtest). TASIT Sarcasm performance correlated significantly with the overall MCCB performance (average of t-scores, r=0.40, p=0.014). Among the 6 subtests, only Speed of Processing survived stepwise modeling of TASIT Sarcasm performance (r=0.45, p=0.006). PSI (used to assess current cognitive abilities) was highly correlated with both MCCB Speed of Processing (r=0.67, p<0.00001) and MCCB overall performance (r=0.63, p=0.00002), demonstrating that PSI was a valid assessment of cognitive abilities in our sample.

TASIT Lies performance also correlated significantly with the overall MCCB performance (average of t-scores, r=0.38, p=0.017). In stepwise modeling of the relationship of the MCCB subtest performance and TASIT lies performance, surviving subtests included Speed of Processing (r_p=0.43, p=0.009) and Working Memory (r_p=0.31, p=0.048).

Supplementary Figures

Supplementary Table 1: ANCOVA results for age-matched subsamples.

Supplementary Table 2: Linear regression results for age-matched subsamples.

Supplementary Figure 1: Visual field overlap with normalized visual features (top row) and face-masked visual features (bottom row) for divergence intervals. SzP viewing deficits were stronger for motion than contrast or luminance. Within faces, however, SzP viewing deficits of the low-level features did not differ significantly. Error bars represent 95% confidence interval. Y axis scaled to roughly equalize non-divergence magnitude and error bars. *$p \leq 0.05$ **$p \leq 0.01$ ***$p \leq 0.001$ ****$p \leq 0.0001$

Supplementary Figure 2: Saccades to faces binned by visual feature intensity. Motion repeated from **Figure 4**. Color scale maximum set to 0.8*(maximum count)/6 in HC map for each visual feature for each group, and that range centered on 0 for the difference map.

Supplementary Figure 3: Saccades to non-face visual features binned by visual feature intensity. Color scale maximum set to 0.8*(maximum count)/6 in HC map for each visual feature for each group, and that range centered on 0 for the difference map.

Supplementary References

Armstrong, K. M., Fitzgerald, J. K., & Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron*, *50*(5), 791–798. http://doi.org/10.1016/j.neuron.2006.05.010

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). Structured Clinical Interview for DSM-IV® Axis I Disorders (SCID-I), Clinician Version, Administration Booklet. American Psychiatric Publishing, Inc.

Gee, A. L., Ipata, A. E., & Goldberg, M. E. (2010). Activity in V4 reflects the direction, but not the latency, of saccades during visual search. *Journal of Neurophysiology*, *104*(4), 2187–2193. http://doi.org/10.1152/jn.00898.2009

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261–276.

Mazer, J. A., & Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron*, *40*(6), 1241–1250.

Russ, B. E., & Leopold, D. A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque., *109*, 84–94. http://doi.org/10.1016/j.neuroimage.2015.01.012

Squire, R. F., Noudoost, B., Schafer, R. J., & Moore, T. (2013). Prefrontal contributions to visual selective attention. *Annual Review of Neuroscience*, *36*, 451–466. http://doi.org/10.1146/annurev-neuro-062111-150439

Zhu, X., & Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. *Cvpr*.

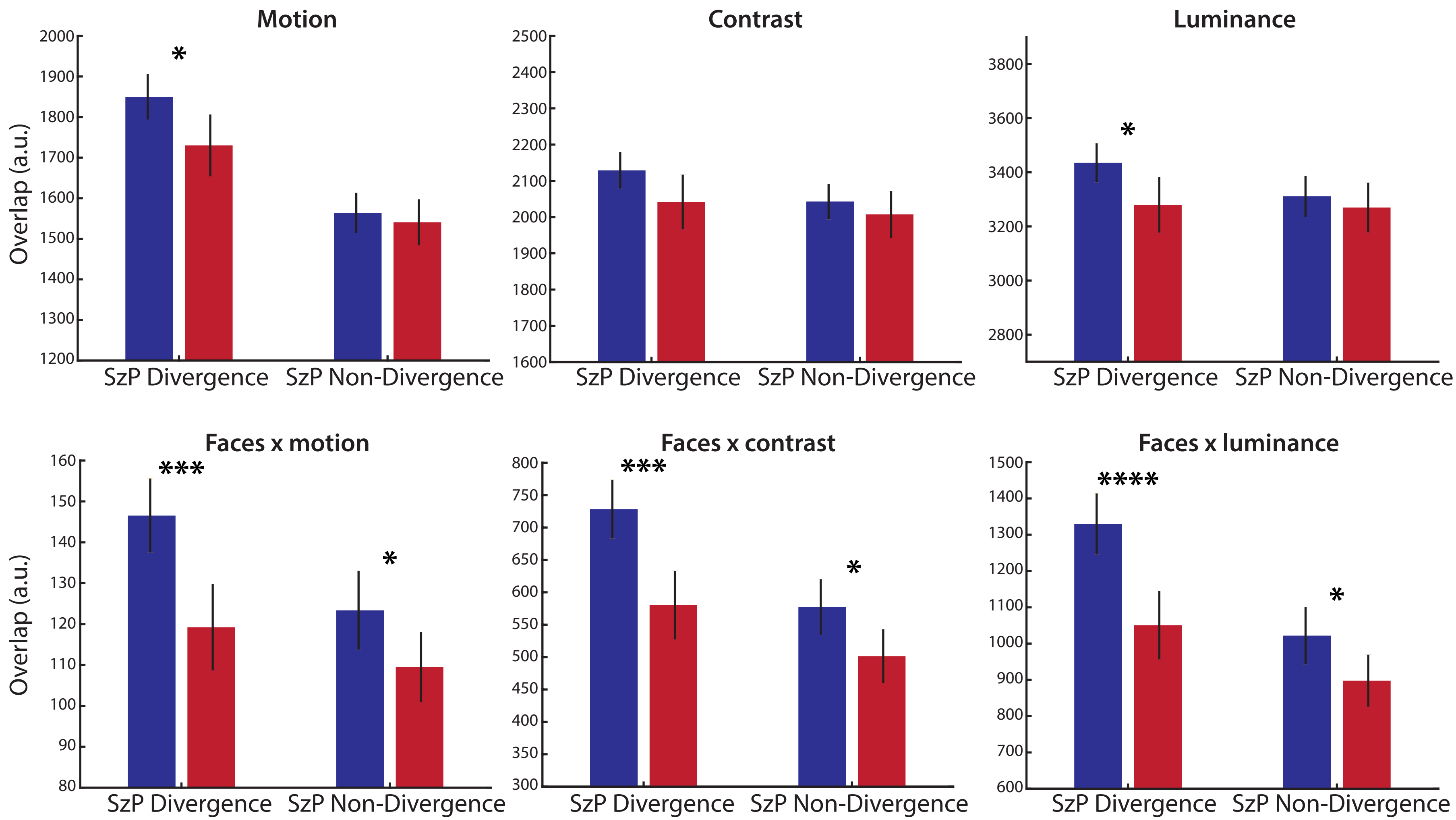
Factor	d.f.	F	p
Visual Scanning ~ TASIT Sarcasm			
Group	1	24.9	0.000
X	1	2.3	0.138
Group*X	1	4.9	0.032
Error	52		
Visual Scanning + Auditory Sarcasm + PSI ~ TASIT Sarcasm			
Group	1	12.9	0.001
X	1	16.7	0.000
Group*X	1	2.3	0.136
Error	51		

Supplemental Table 1: ANCOVA results after age exclusion

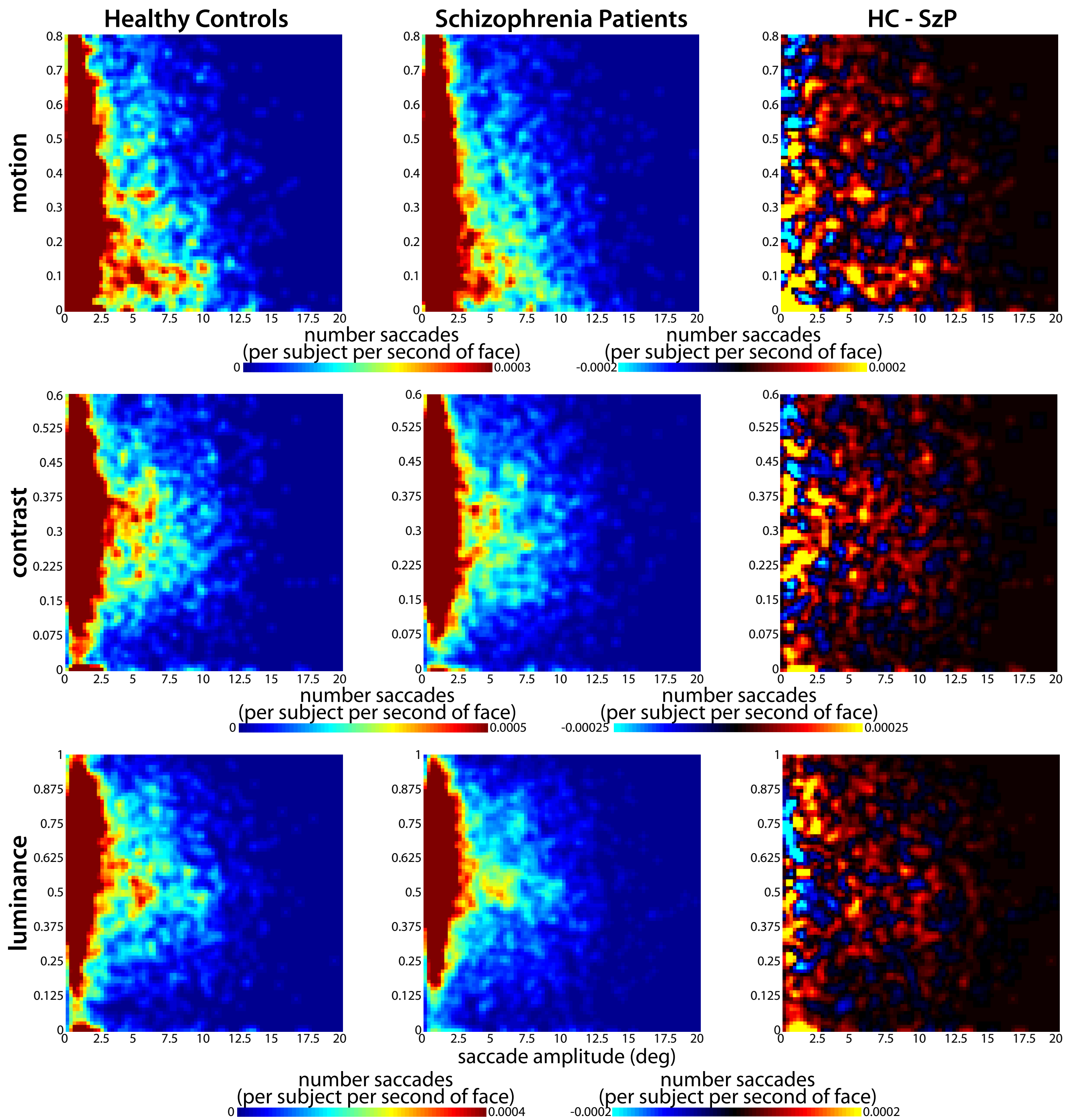
Factor	t-statistic	p
HC model		
Visual Scanning	-2.57	0.019
PSI	0.09	0.928
Auditory Sarcasm	3.05	0.007
Number of observations: 22		Error degrees of freedom: 18
R^2=0.504		
F=6.09	p=0.00479	
SzP model		
Visual Scanning	0.59	0.557
PSI	3.14	0.004
Auditory Sarcasm	0.62	0.542
Number of observations: 33		Error degrees of freedom: 29
R^2=0.31		
F=4.34	p=0.0121	

Supplemental Table 2: Within group multivariate models after age exclusion (Visual Scanning + Auditory Sarcasm + PSI ~ TASIT Sarcasm)

Supplemental Figure 1



Supplemental Figure 2



Supplemental Figure 3