

Original Article

Cite this article: Patel GH *et al* (2021). What you see is what you get: visual scanning failures of naturalistic social scenes in schizophrenia. *Psychological Medicine* **51**, 2923–2932. <https://doi.org/10.1017/S0033291720001646>

Received: 24 July 2019

Revised: 25 March 2020

Accepted: 12 May 2020

First published online: 5 June 2020

Key words:


Attention; face emotion recognition; motion; social cognition; visual search

Author for correspondence:

Gaurav H. Patel,

E-mail: gph2114@cumc.columbia.edu

What you see is what you get: visual scanning failures of naturalistic social scenes in schizophrenia

Gaurav H. Patel^{1,2} , Sophie C. Arkin³, Daniel R. Ruiz-Betancourt¹, Heloise M. DeBaun¹, Nicole E. Strauss⁴, Laura P. Bartel^{1,2}, Jack Grinband^{1,2}, Antígona Martínez^{1,5}, Rebecca A. Berman⁶, David A. Leopold⁶ and Daniel C. Javitt^{1,2,5}

¹Department of Psychiatry, Columbia University Irving Medical Center, New York, USA; ²Division of Experimental Therapeutics, New York State Psychiatric Institute, New York, USA; ³Department of Psychology, University of California Los Angeles, Los Angeles, USA; ⁴Palo Alto University, Palo Alto, USA; ⁵Schizophrenia Research Division, Nathan Kline Institute for Psychiatric Research, Orangeburg, USA and ⁶Section on Cognitive Neurophysiology and Imaging, National Institute for Mental Health, Bethesda, USA

Abstract

Background. Impairments in social cognition contribute significantly to disability in schizophrenia patients (SzP). Perception of facial expressions is critical for social cognition. Intact perception requires an individual to visually scan a complex dynamic social scene for transiently moving facial expressions that may be relevant for understanding the scene. The relationship of visual scanning for these facial expressions and social cognition remains unknown. **Methods.** In 39 SzP and 27 healthy controls (HC), we used eye-tracking to examine the relationship between performance on The Awareness of Social Inference Test (TASIT), which tests social cognition using naturalistic video clips of social situations, and visual scanning, measuring each individual's relative to the mean of HC. We then examined the relationship of visual scanning to the specific visual features (motion, contrast, luminance, faces) within the video clips.

Results. TASIT performance was significantly impaired in SzP for trials involving sarcasm ($p < 10^{-5}$). Visual scanning was significantly more variable in SzP than HC ($p < 10^{-6}$), and predicted TASIT performance in HC ($p = 0.02$) but not SzP ($p = 0.91$), differing significantly between groups ($p = 0.04$). During the visual scanning, SzP were less likely to be viewing faces ($p = 0.0001$) and less likely to saccade to facial motion in peripheral vision ($p = 0.008$).

Conclusions. SzP show highly significant deficits in the use of visual scanning of naturalistic social scenes to inform social cognition. Alterations in visual scanning patterns may originate from impaired processing of facial motion within peripheral vision. Overall, these results highlight the utility of naturalistic stimuli in the study of social cognition deficits in schizophrenia.

Introduction

Deficits in social cognition are a major cause of psychosocial disability in schizophrenia, but the underlying mechanisms remain incompletely understood (Green, Horan, & Lee, 2015, 2019). In addition to the perception of auditory cues and cognitive processing speed, visual scanning of social scenes – the exploration of the scenes with saccadic eye-movements – is an important component of social cognition (Green, Horan, & Lee, 2015; Zaki & Ochsner, 2009). Visual scanning patterns are driven in part by variation in low-level visual features, such as luminance, contrast, and motion speed (Kusunoki, Gottlieb, & Goldberg, 2000; Marsman *et al.*, 2016; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010; White *et al.*, 2017), as well as by as-yet unexplained cognitive factors (Wilming *et al.*, 2017). Disturbances in the processing of all of these visual features have been reported previously in schizophrenia patients (SzP) (Butler *et al.*, 2009; Calderone *et al.*, 2013; Chen *et al.*, 1999; Javitt, 2009; Martínez *et al.*, 2018; Taylor *et al.*, 2012) and have been linked to social cognitive deficits (Green *et al.*, 2015, 2019; Javitt, 2009). However, how processing of visual features affects visual scanning and in turn social cognition in SzP has not been studied.

To examine the relationship between visual processing, visual scanning, and social cognition, we used a validated naturalistic test of social cognition – The Awareness of Social Inference Test (TASIT) (McDonald, Flanagan, Rollins, & Kinch, 2003) – while tracking eye-movements. Unlike the static stimuli in traditional tests of social cognition, such as ER-40 or Reading the Mind in the Eyes Test (Pinkham, Penn, Green, & Harvey, 2015), TASIT contains the motion dynamics of real-world social situations, which are the strongest predictors of visual scanning in naturalistic scenarios (White *et al.*, 2017).

The TASIT consists of a series of video-based vignettes in which 2–3 individuals interact. In each vignette, the main actor is either being sarcastic or lying to another character. In the sarcasm videos, the main actor uses exaggerated facial expressions and auditory prosody to indicate that the intended meaning is counterfactual to the plain meaning of his utterance. Correctly answering the questions about the sarcasm videos, then, requires the viewer to optimally detect both the visual and auditory social cues. By contrast, lies are detected by comparing the content of what the actor is saying to information conveyed elsewhere in the video, so that auditory prosody and facial expression are not critical factors.

SzP show reliable deficits in this task and are particularly impaired in the detection of sarcasm (Pinkham et al., 2015; Sparks, McDonald, Lino, O'Donnell, & Green, 2010). Furthermore, TASIT deficits correlate significantly with real-world social functioning, supporting its ecological relevance (Pinkham et al., 2015). This test thus serves as a potentially powerful platform to use for investigating underlying mechanisms related to social cognition. We hypothesized that visual scanning patterns would differ in SzP compared to HC, and that visual scanning would predict TASIT performance in both groups independent of other relevant factors, such as the detection of sarcasm in spoken sentences, recognition of facial expressions, and cognitive processing speed (Holdnack, Prifitera, Weiss, & Saklofske, 2015).

During visual scanning, each saccade represents a decision to move the eyes from the current location to another visual feature that may provide additional information about the social scene (Corbetta, Patel, & Shulman, 2008; Patel, Sestieri, & Corbetta, 2019). Given that faces are a key visual feature used to make inferences about the mental states of the people in the scene, we hypothesized that the divergent visual scanning patterns in SzP would result in decreased viewing of faces.

Sometimes the faces are near the current focus (in central vision, $<5^\circ$ from the fovea or center of the retina) and sometimes they are further away in peripheral vision ($>5^\circ$ from the fovea). In general, SzP are impaired in the processing of low-level visual features that depend on magnocellular visual processing pathways, in particular motion (Butler et al., 2005; Javitt, 2009; Martinez et al., 2018). Facial expressions in the real world involve slow subtle movements (Sowden, Schuster, & Cook, 2019), and SzP seem particularly impaired in processing moving facial expressions (Arnold, Iaria, & Goghari, 2016; Johnston et al., 2010; Kohler et al., 2008). Since magnocellular processing dominates peripheral vision, SzP may have greater deficits in the processing of peripheral *v.* central moving facial expressions (Dias et al., 2020; Javitt, 2009; Martinez et al., 2018). Therefore, we examined the visual scanning of faces in peripheral vision relative to low-level visual features (motion, contrast, and luminance), hypothesizing that the likelihood that SzP make saccades to moving facial expressions in peripheral vision would be reduced compared to healthy controls (HC).

However, methods for comparing visual scanning patterns between groups and linking them to the visual features in a video largely do not exist, thus requiring us to adapt or create a number of analytical techniques for this study. A key principle underlying our approach was to use the HC visual scanning pattern as the 'gold standard' by which to compare SzP patterns, and then to examine the intervals during which SzP visual scanning patterns diverged from the HC. This approach takes advantage of previous observations that HC show highly convergent scan patterns of naturalistic scenes despite the lack of explicit instruction where to look (Marsman et al., 2016; Wilming et al., 2017).

We then used neurophysiologically based models of the visual field and saccade generation to examine the relationship of what visual features fell into central *v.* peripheral vision and visual scanning. Combined with automated detection and classification of visual features, these techniques allowed us to model the processing of the visual scene for each individual in a way that mimics their experience of the visual scanning of real-life social situations, thus allowing us to test our hypotheses about the link between divergent visual scanning patterns and social cognition.

Materials and methods

Participants

Forty-two SzP and 30 HC were recruited with informed consent in accordance with New York State Psychiatric Institute's Institutional Review Board. Three SzP and three HC were excluded from further behavioral analyses because of problems with the eye-tracking data, such as gross errors in calibration, leaving 39 SzP and 27 HC in the analyses. SzP were moderately ill, domiciled, recruited from the community, and stabilized on medication. HC were demographically matched to SzP with no history of major psychiatric disorders. See online Supplementary Materials for more details.

Behavioral testing

In multiple separate sessions, participants were evaluated on a combination of standard neuropsychological assessments, symptomology scales, and other behavioral tests, including the processing speed index (PSI) in the WAIS-III (Wechsler, 1997), Attitudinal Prosody (auditory sarcasm) (Leitman, Ziwich, Pasternak, & Javitt, 2006; Orbelo, Grim, Talbott, & Ross, 2005), Penn Emotion Recognition (ER-40) (Heimberg, Gur, Erwin, Shtasel, & Gur, 1992), and TASIT (McDonald et al., 2003) with eye-tracking. PSI was chosen to represent cognitive processing speed as it best reflects full-scale cognitive capabilities in SzP and consists of visual search tasks that require visual scanning (Bulzacka et al., 2016). SzP also performed the MATRICS Cognitive Consensus Battery (MCCB) (Nuechterlein et al., 2008).

Group comparisons and correlations

All group and condition comparisons (except for those detailed in *Saccades to Visual Features* below) were performed using repeated-measures ANOVAs and post-hoc *t* tests with a threshold of $p < 0.05$. All correlations between conditions were performed using linear regression, with group co-variance assessed with ANCOVAs. To assess the relative contributions of the various neuropsychological measures to predicting TASIT performance, the *z*-transformed measures were entered into a linear regression model predicting TASIT performance (also *z*-transformed) to give their partial correlations. These relative weights were then used to create a composite score to serve as a univariate predictor of TASIT performance. Since HC performance was near the maximum for TASIT, TASIT performance was arcsin transformed to perform Normal statistics.

Visual scanning measures and analyses

To compare variability in visual scanning patterns between groups, we first calculated the mean eye-position and the standard

deviation (s.d.) elliptical boundary on each video frame for each group. The elliptical boundary axes were determined by s.d. of the mean x and the mean y positions. We then compared group differences in the ellipse area averaged across frames, correcting for autocorrelation ($\text{number_of_frames}^* = \text{number_of_frames}/2T_e$, where T_e is the number of frames the autocorrelation takes to drop to $1/e$). We also counted, for each group, the percentage of individuals that fell outside of the HC 2 s.d. elliptical boundary. Each individual HC's eye position was compared to the leave-one-out average and 2 s.d. elliptical boundary derived from all of the other HC.

To assess the degree of divergence of each individual's eye position from the mean HC position, we calculated the z -transformed distance for each individual for each video frame. The z -transformed distance is the distance between the individual's eye position and the mean HC position divided by the standard deviation of the HC eye position distribution (again individual HC were compared against a leave-one-out average of the other HC). This measure (log transformed for Normal statistics) emphasizes divergence by more heavily weighting eye positions that are distant from the HC mean during intervals when HC eye positions are relatively convergent, or when the HC standard deviation is small.

Visual features within the visual field

To quantify which visual features each individual looked at, we simulated the visual processing of the video frame using algorithms designed to mimic the processing by visual cortex. We first applied automated detection and outlining of faces to generate binary face masks (Zhu & Ramanan, 2012) (Fig. 3a). Then each video frame was divided into $1^\circ \times 1^\circ$ cells and the strength of low-level visual features (motion speed, contrast, and luminance) was quantified for each cell in each video frame (Russ & Leopold, 2015) (Fig. 3b). These visual features were then temporally smoothed and normalized to the maximum intensity to mimic processing in the visual cortex, and the low-level visual features were multiplied by the face masks to generate maps of visual features within faces.

To determine which visual features each participant was seeing, we modeled the visual field with a 2D representation of the V4 cortical magnification factor (Serenio et al., 1995) centered on the eye position for each video frame and smoothly weighted by its proximity to the center of the visual field (Fig. 3a). This visual field model was then multiplied by each visual feature map to generate a salience map (Fig. 3c). The surface integral of this salience map summarizes how much each visual feature is represented in the visual field for that participant on each video frame. To compare what each group was viewing, these summary measures were averaged across frames and individuals. See online Supplementary Materials for further details.

Saccades to visual features

We next examined the number of saccades made to faces as a function of saccade amplitude in each group. For every saccade made to a face, saccade amplitude was binned by 0.25° intervals for each individual across sarcasm or lie trials before group averaging. To then examine which visual features (motion, contrast, luminance) drove saccades to faces in peripheral vision, we quantified the intensity of the visual feature within that face in the 133 ms interval prior to the saccade. We then plotted the density

of saccades as a function of both saccade amplitude and visual feature strength for each group, and then searched the group difference map for significant clusters of saccades. Saccade amplitudes $\geq 5^\circ$ were defined as saccades to peripheral faces, as that amplitude approximates the boundary between foveal and peripheral processing (Nieuwenhuys, Voogd, & van Huijzen, 2008) (see online Supplementary Materials for details). False-positive rates were determined by permutation testing, shuffling group labels, and repeating the analysis 10 000 times.

Results

TASIT performance

Demographically, SzP were matched to HC, with a small increase in mean age [40.6(11.0) *v.* 35.2(9.3), $p = 0.04$] in SzP (Table 1). As predicted, SzP were highly impaired in the comprehension of TASIT clips (group: $F_{1,64} = 25.63$, $p < 10^{-5}$, Cohen's $d = 1.3$), with a greater impairment in the comprehension of sarcasm *v.* lies (group \times sarcasm/lie: $F_{1,64} = 9.43$, $p = 0.003$, Fig. 1a).

Eye position variability and visual scanning performance

Across trials, SzP as a group made more saccades than HC [1009.1(242.1) *v.* 970.8(233.2), $t_{15} = 4.6$, $p = 0.0003$] with no difference between sarcasm and lie trials, resulting in a lower mean fixation duration of 536 *v.* 560 ms. SzP eye positions overall were also more variable across all TASIT video frames compared to HC: the average area of the ellipse representing the standard deviation of the x and y eye position across video frames was 32.5% larger in SzP *v.* HC ($t_{880} = 5.0$, $p < 10^{-6}$, ellipses in Fig. 1b).

We further quantified visual scanning variability of each group as the percentage of participants that fell outside the HC 2 s.d. elliptical boundary on each video frame. SzP visual scanning variability was substantial, with an average of 18.7(10.7)% of eye positions more than 2 s.d. from the HC mean eye position, while the HC eye positions rarely deviated [6.1(2.3)%, consistent with a Normal distribution]. Reversing the analysis (comparing all participants to the SzP mean eye position) resulted in 8.7(3.2)% of HC eye positions falling outside the SzP 2 s.d. elliptical boundary compared to SzP [6.2(5.2)%], further supporting increased SzP eye position variability. Moreover, the SzP visual scanning variability was not evenly distributed through all frames of the videos: there were 78 intervals across the 16 videos during which ³50% of SzP were outside of the 2 s.d. ellipse.

SzP visual scanning divergence

To quantify individual visual scanning divergence, we calculated the z -transformed distance of each individual's eye position *v.* the mean HC position. Across video frames, the SzP mean z -transformed distance of 1.57(0.54) was significantly different from the HC mean z -transformed distance of 1.31(0.16), and did not differ for lie *v.* sarcasm trials (ANOVA, group: $F_{2,64} = 5.8$, $p = 0.02$; lie/sarcasm: $F_{2,64} = 0.05$, $p = 0.83$; group \times lie/sarcasm: $F_{2,64} = 0.2$, $p = 0.7$). Similar to the variability measure above, SzP were not divergent on every video frame. Rather, there were troughs where SzP and HC divergence was similar, and peaks where SzP were much more divergent than HC (Fig. 1c). Applying a divergence threshold of a z -transformed distance of 2 (black dashed line Fig. 1c), we found that SzP diverged on 11.8% of the total frames in 320 intervals.

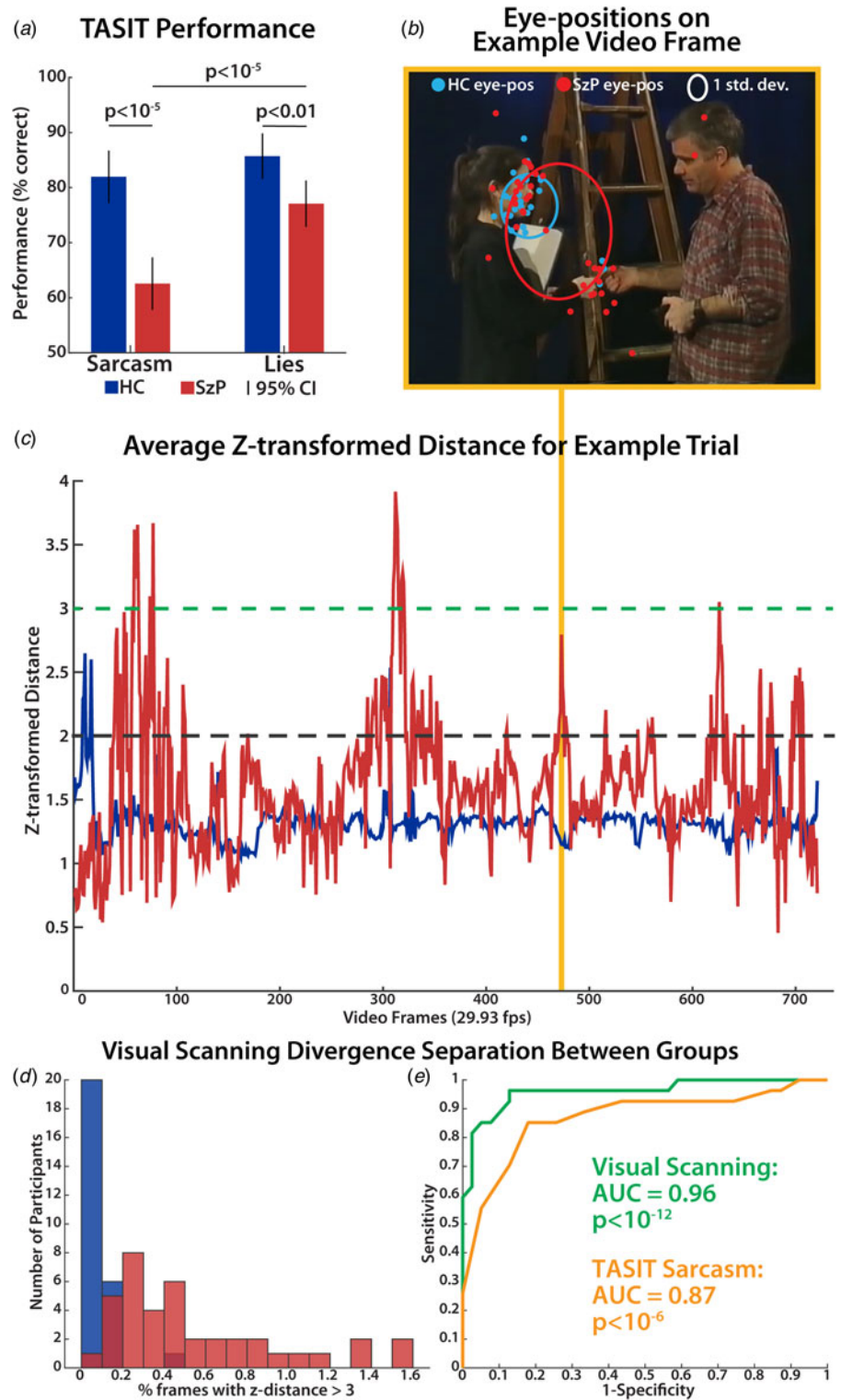


Fig. 1. TASIT and visual scanning performance. (a) TASIT performance by group where 50% is chance. (b) Example video frame of gaze positions of HC (blue) and SzP (red). HC eye positions in blue are tightly clustered v. the scatter of the red eye positions of the SzP, as demonstrated by the difference in the area of the elliptical boundary representing 1 standard deviation for each group. (c) SzP and HC visual scanning divergence time-course. Divergence measured as average z-transformed distance. Dotted lines mark thresholds for various analyses: green for ROC analyses (Figs 1d and e) and black for visual feature analyses (Fig. 3). (d) Histogram showing the distribution of participants in each group of % video frames with eye position z-transformed distance >3 . (e) ROC curves for group segregation based on visual scanning (z-transformed distance >3) and TASIT.

We then used an ROC analysis to examine whether these SzP divergence peaks were driven by individual outliers or by a systematic deviation across the group. With a threshold of z-transformed distance >2 , SzP could be separated from HC with an ROC AUC = 0.83. This AUC was greater than expected by chance ($p = 0.002$) even given the biasing nature of the analysis, demonstrating that these peaks in SzP divergence were driven by a systematic deviation away from the HC visual

scanning pattern during these intervals. Increasing the threshold to z-transformed distance >3 (green dashed line Fig. 1c) generates an ROC AUC = 0.96 ($p < 0.0001$), substantially better than TASIT sarcasm performance itself (AUC = 0.87, $p = 0.05$, Figs 1d and e). These results demonstrate the potential of visual scanning as a diagnostic biomarker: if an individual's eye position is >3 z-transformed distance away from the HC mean for more than 0.55 s (0.2%) of the 276 s of TASIT

Table 1. Demographics

Demographic	SzP	HC	Statistics
Age	40.6(11.0)	35.2(9.3)	$t(61) = 2.1, p = 0.04$
Patient SES	30.3(13.5)	34.4(13.5)	$t(56) = 1.2, NS$
Parent SES	39.7(15.4)	45.3(13.5)	$t(59) = 1.5, NS$
Gender (F/total)	10/39	11/27	
Edinburgh Inventory	17.3(3.9)	17.5(3.4)	$t(59) = 0.2, NS$
Vision (40in) (median)	20/32	20/32	Wilcoxon RS $p = 0.7$
Color blindness (# subjects)	0/39	0/27	
PANSS Positive	15.8(6.1)		
PANSS Negative	13.3(3.6)		
PANSS Total	56.5(14.5)		
Antipsychotic dose (CPZ equivalents)	481(671)		

The bold text marks the significant results.

sarcasm video clips, in our sample, they are highly likely to have schizophrenia.

Visual scanning v. TASIT performance

We next examined the relationship of visual scanning and TASIT performance (Fig. 2). The slope of the relationship between scanning behavior and TASIT performance (arcsin transformed for Normal statistics) differed significantly between groups (group \times z-distance: $F_{2,64} = 4.6, p = 0.04$), with a significant relationship observed in HC ($r = 0.46, p = 0.02$) but not SzP ($r = 0.02, p = 0.91$), indicating that visual scanning performance was related to the comprehension of the TASIT sarcasm videos only in HC (Fig. 2a). Visual scanning performance did not correlate with TASIT lie performance either within or between groups.

To understand the lack of correlation in SzP, we next explored the relationship of visual scanning and TASIT sarcasm performance relative to measures of other abilities critical for following these social situations: the detection of auditory sarcasm, the speed of cognitive processing (measured by PSI), and face-emotion recognition (measured by ER-40). SzP were significantly impaired in all of these measures (Table 2) with a significant correlation with TASIT sarcasm performance across groups (Table 3). For auditory sarcasm, the group \times covariate interaction was not significant, but similar to visual scanning, the correlation was significant in HC ($r = 0.55, p = 0.003$) but not SzP ($r = -0.25, p = 0.12$) (Fig. 2b). Cognitive processing speed did exhibit a significant group \times covariate interaction but with the opposite pattern as visual scanning and auditory sarcasm, with a strong correlation in SzP ($r = 0.50, p = 0.001$) but not HC ($r = 0.08, p = 0.68$) (Fig. 2c). Face-emotion recognition did not exhibit a group \times covariate interaction nor a significant correlation within either group (HC: $r = 0.20, p = 0.31$; SzP: $r = 0.30, p = 0.06$).

We next examined which combinations of these measures independently predicted TASIT performance in each group and across groups. For HC, only visual scanning and auditory sarcasm remained significant in stepwise regression (visual scanning: $r_p = 0.38, p = 0.021$; auditory sarcasm: $r_p = 0.48, p = 0.005$) (Fig. 2d).

For SzP, only cognitive processing speed remained significant. Combining the three within-group significant predictors of TASIT sarcasm performance (visual scanning, auditory sarcasm, and cognitive processing speed) explained 42% of the variance in TASIT performance across both groups, or 24% of the variance in TASIT performance after accounting for group membership (composite score: $F_{2,64} = 19.1, p = 0.00005$, group: $F_{2,64} = 13.2, p = 0.0006$), with no difference in the relationship between groups (group \times composite score: $F_{2,64} = 0.9, p = 0.34$; Fig. 2e). Exploration of the other relationships in SzP revealed a negative correlation between visual scanning and auditory sarcasm ($r = -0.40, p = 0.01$) not present in HC ($r = 0.16, p = 0.4$), as well as a positive correlation of face-emotion recognition and cognitive processing speed in SzP ($r = 0.35, p = 0.033$) but not HC ($r = 0.06, p = 0.8$, Fig. 2d).

TASIT sarcasm performance in SzP correlated with the MCCB composite score ($r = 0.40, p = 0.014$), though within MCCB only the Speed of Processing (SoP) domain score ($r = 0.45, p = 0.006$) was significant. TASIT sarcasm performance in SzP also correlated strongly with the Brief Assessment of Cognition in Schizophrenia Symbol Coding score (BACS-SC, $r = 0.43, p = 0.007$), but not category fluency ($r = 0.24, p = 0.16$). Antipsychotic medication dose did not correlate with TASIT performance and controlling for age did not change the above results (details in online Supplementary Materials).

Visual features missed by SzP

We next examined what visual features SzP may be missing compared to HC during the SzP peak divergence intervals (Figs 3a–c for analysis details). A strong group \times interval effect ($F_{2,64} = 40.2, p < 10^{-7}$) demonstrated a 20% decrease in the amount of time SzP spent looking at faces during the SzP divergence intervals compared to HC ($t_{64} = 4.1, p = 0.0001$, Fig. 3d). However, for basic visual features (motion, contrast, and luminance), the SzP deficit was much weaker, with a somewhat larger effect for motion *v.* the other visual features. A group \times interval \times visual feature effect ($F_{6,64} = 3.2, p = 0.04$) showed that SzP spent a little less time looking at motion (6.5%) during the SzP divergence intervals compared to contrast and luminance (4.1% and 4.5%, respectively, Fig. 3e and online Supplementary Fig. S1).

We also did not find significant differences in the viewing of basic visual features *within* faces (Fig. 3f and online Supplementary Fig. S1): there were no group \times interval \times visual feature effects ($F_{6,64} = 2.3, p = 0.11$). In addition, the 18.6–21.0% decrease in the viewing of the three face-masked visual features more closely resembled the 20% decrease in the viewing of faces than to the 4.1–6.5% decrease in viewing the three non-face-masked visual features discussed above, suggesting that the deficit was more specific to viewing faces than any of the basic visual features.

Visual features driving divergent visual scanning patterns

Finally, we examined what visual features may be driving the divergence in the visual scanning of the TASIT videos in the SzP. Both groups made more saccades in sarcasm trials *v.* lie trials (trial type: $F_{2,64} = 28.2, p = 10^{-6}$) with no significant difference between groups. In total, 18.8% of saccades to faces were made to faces in the periphery ($>5^\circ$, or more than five boxes in Fig. 3b). Of those saccades, HC consistently made more saccades to faces in the 5–7.5° range than SzP (permutation testing: 10/10 bins, $p = 0.037$, green zone in Fig. 4a).

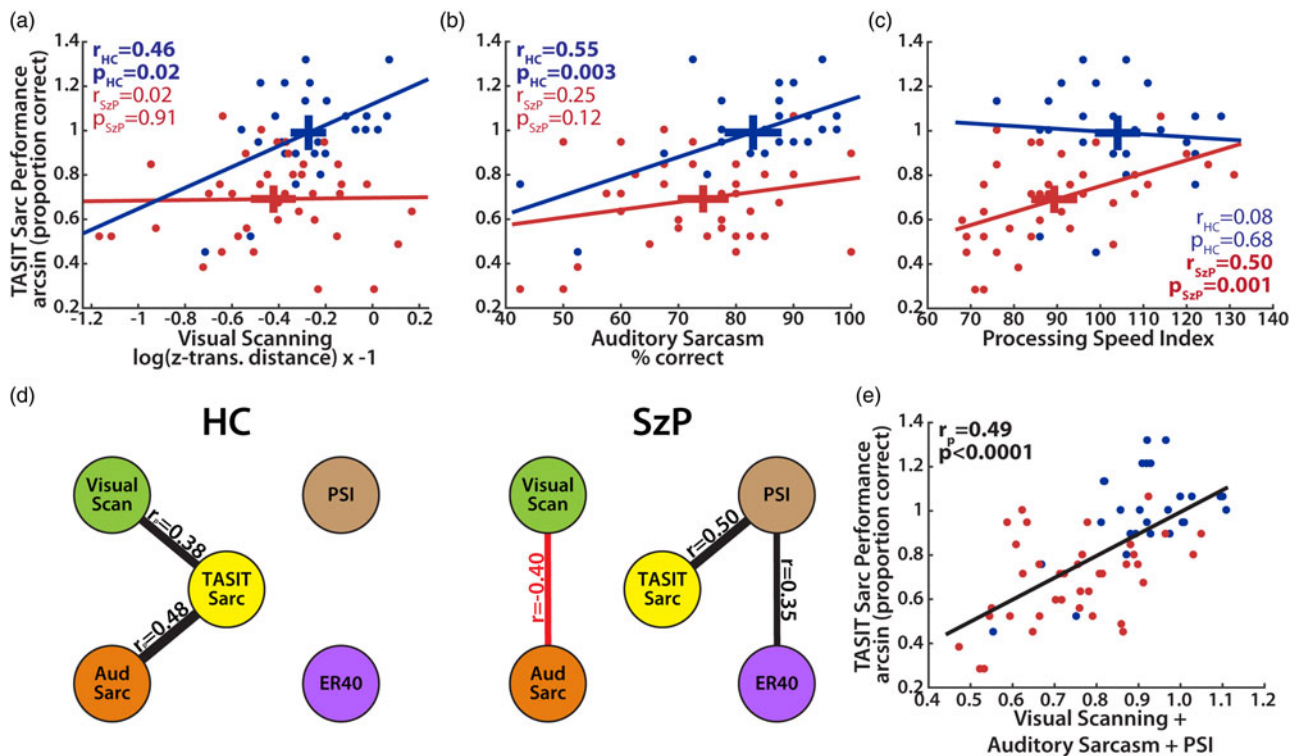


Fig. 2. (a) Correlation of visual scanning with TASIT sarcasm performance. Crosses represent horizontal and vertical 95% confidence intervals. Visual scanning is measured as z-transformed distance, log transformed for Normal statistics, and inverted so that higher values represent better performance. (b) Correlation of auditory sarcasm detection performance and TASIT sarcasm performance by group. (c) Correlation of cognitive processing speed as measured by the WAIS-IV processing speed index (PSI) by group. (d) Summary of correlations between TASIT sarcasm, visual scanning auditory sarcasm, cognitive processing speed (PSI), and face-emotion recognition (measured by ER-40) in HC and SzP. (e) Correlation of composite score representing the combination of visual scanning, auditory sarcasm, and PSI with TASIT sarcasm performance.

We then quantified the relationship between the amplitude of the saccade to faces and the visual features present in the faces before the saccade. The number (density) of saccades to faces in the peripheral visual field made by each group were stratified by the strength of the low-level visual features in those faces (Figs 4b–d and online Supplementary Fig. S2). For motion, the saccade density plots showed a high-density cluster of saccades in HC to slow-motion speed starting in the 5–7.5° saccade amplitude range and extending to 10° (orange circle in Fig. 4b). This range of speeds (<0.4 pixels/frame) matched that of facial expressions (Sowden *et al.*, 2019). There was no similar cluster in SzP (orange circle in Fig. 4c), leading to a significant group difference (green circle in Fig. 4d, $p = 0.008$, permutation testing). No clusters survived significance in the saccade density plots for other visual features or for saccade density plots to non-face locations, indicating that the greater number of saccades to peripheral faces in HC was driven specifically by face motion as opposed to other types of motion (online Supplementary Fig. S3).

Discussion

To our knowledge, this is the first study to evaluate the role of visual scanning in the comprehension of dynamic naturalistic social scenes in SzP. There were five main findings. First, as expected, SzP were more impaired in the comprehension of the TASIT sarcasm *v.* lie videos. Second, SzP visual scanning patterns often diverged from those of HC. Third, visual scanning integrity and

auditory sarcasm detection predicted TASIT sarcasm performance in HC, whereas only cognitive processing speed predicted TASIT performance in SzP. Fourth, SzP often missed viewing the faces that drew the gaze of HC. Finally, SzP orient less often to moving facial expressions in the periphery than HC. Overall, these findings suggest that SzP are unable to rely on the detection of moving facial expressions in the periphery to efficiently guide visual scanning of complex dynamic social scenes as HC do, instead relying on alternate strategies that rely on cognitive abilities to, at least partially, overcome the social cognition deficits.

TASIT broadly tests the ability to use auditory and visual social cues to make inferences about the actors' mental states, and correlates with overall social functioning (Pinkham *et al.*, 2015). Despite the fact that detection of both lies and sarcasm requires inference of the speaker's internal mental state, different types of information are used to make the inference in the two situations. For both lies and sarcasm, the viewer must understand that the communicated information is counterfactual to reality. However, in the case of lies (as portrayed in the TASIT), the information is communicated by comparing the information content of what the main actor is saying at different points in the video.

By contrast, in the case of sarcasm, the information is also communicated through modulation of the tone of voice [attitudinal prosody (Leitman *et al.*, 2006, 2010)] and facial expressions. The TASIT was normed to be equally sensitive to impairments in sarcasm and lie detection in traumatic brain injury (McDonald *et al.*, 2006). The differential deficit in sarcasm *v.* lie detection therefore suggests differential impairment in the ability to utilize

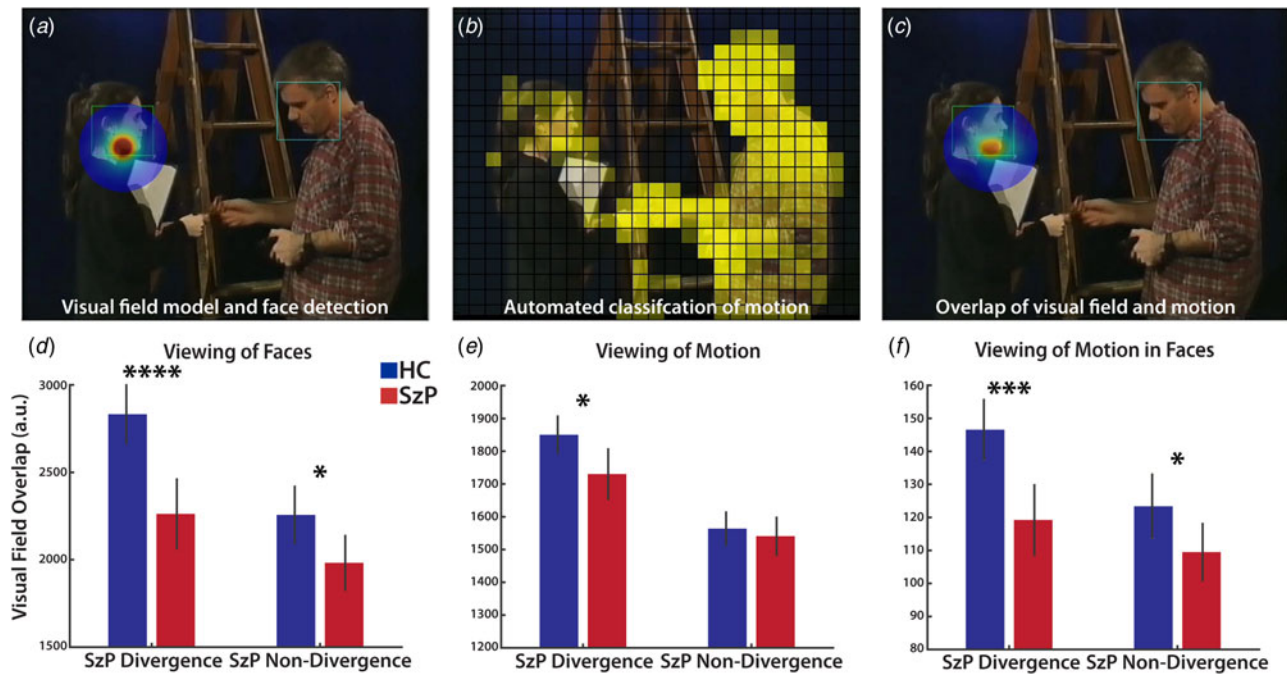


Fig. 3. Viewing of visual features during intervals when SzP eye positions are divergent. (a) Modeling of where in the video frame each participant was looking by centering a visual field model on their eye position (rainbow halo). Green squares represent automated face detection boundaries. (b) Example visual feature map for motion speed, normalized by maximum motion and averaged over previous 133 ms interval. Each visual feature cell was approximately $1^\circ \times 1^\circ$, with the size rounded to represent each video frame as 25×21 cells. (c) The motion map (b) was filtered through the visual field model (a), creating a map of which visual feature the participant was viewing on each frame, weighted by distance from the center of the visual field. Face-masked visual feature maps (f) masked the non-normalized versions of these visual field feature maps with a face mask (square outlines). (d) Faces viewed during SzP divergence v. non-divergence intervals (SzP mean z-transformed distance >2 or <-2 , respectively; black dotted line in Fig. 1c). The ‘visual field overlap’ measure represents the combination of the time spent viewing a visual feature and the proximity of that visual feature to the center of the visual field. The greater viewing of faces by HC v. SzP during the divergence intervals demonstrates that HC were often viewing faces during these intervals and SzP were not. (e) Viewing of motion during SzP divergence and non-divergence intervals. (f) Viewing of motion within faces during SzP divergence and non-divergence intervals. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

sensory information and to orient to critical features of the environment. The specific deficit in the sarcasm v. lie trials also indicates that performance deficits were not caused solely by reduced vigilance due to medications or other reasons, and that SzP were generally able to both follow the dialogue and actions on screen and understand the questions.

The systematic divergence of SzP and HC visual scanning patterns suggests that SzP miss seeing certain visual features that HC spontaneously decide are important, namely faces. The naturalistic stimuli allowed us to explain why SzP may not have been looking at those faces. The failure to process motion (and specifically biological motion) in central vision is not a new finding (Butler et al., 2005; Chen, Levy, Sheremata, & Holzman, 2004; Martinez et al., 2018; Okruszek & Pilecka, 2017). However, our findings suggest that motion processing deficits in SzP may have the largest impact in peripheral vision. These peripheral deficits also appear to be independent of previously detailed deficits in

face-emotion recognition deficits, usually measured in central vision (Arnold et al., 2016; Johnston et al., 2010; Kohler et al., 2008). Our findings suggest that the deficits in motion processing in peripheral vision reduce the likelihood that faces are ever brought into central vision for further inspection, and when they are, the previously detailed face-emotion recognition deficits may additionally impact the understanding of the social scene.

The correlation of visual scanning and TASIT sarcasm performance in HC but not SzP suggests that most SzP are unable to rely on stimulus-driven visual scanning to locate these faces. The ones that are able to mimic the HC visual scanning pattern face another problem – they are unable to detect auditory expressions of sarcasm, a negative correlation that suggests that SzP can have intact visual or auditory processing but not both. Instead, the better-performing SzP may be using an alternative strategy, as evidenced by the relationship of TASIT performance with cognitive processing speed in SzP, the correlation between face-emotion recognition and cognitive processing speed, and the specificity of TASIT sarcasm performance correlating with BACS-SC (which requires multiple fast saccades between pre-specified locations) and not category fluency (which does not require eye movements). Rather than the ‘bottom-up’ stimulus-driven strategy employed by the HC, these SzP may be using a ‘top-down’ strategy of targeting saccades to the known locations of the faces and then making fast decisions about the expressions in those faces. This alternative strategy may reflect either compensation or inversely a second deficit in the worst-performing SzP.

Table 2. SzP v. HC performance

Measure	t_{64}	p
Visual scanning	-2.3	0.027
Auditory sarcasm	-2.6	0.012
General cognition	-3.8	0.0004
Face-emotion recognition	-3.3	0.001

Table 3. Predictors of TASIT Sarcasm performance

Measure	Group	Covariate	Group × covariate
Visual scanning	$F_{2,64} = 32.3, p < 10^{-6}$	$F_{2,64} = 1.7, p = 0.2$	$F_{2,64} = 4.6, p = 0.036$
Auditory sarcasm	$F_{2,64} = 27.6, p < 10^{-5}$	$F_{2,64} = 10.1, p = 0.002$	$F_{2,64} = 2.1, p = 0.15$
General cognition	$F_{2,64} = 23.7, p < 10^{-5}$	$F_{2,64} = 5.6, p = 0.021$	$F_{2,64} = 4.8, p = 0.031$
Face-emotion recognition	$F_{2,64} = 24.0, p < 10^{-5}$	$F_{2,64} = 4.6, p = 0.036$	$F_{2,64} = 0.1, p = 0.8$

The bold text marks the significant results.

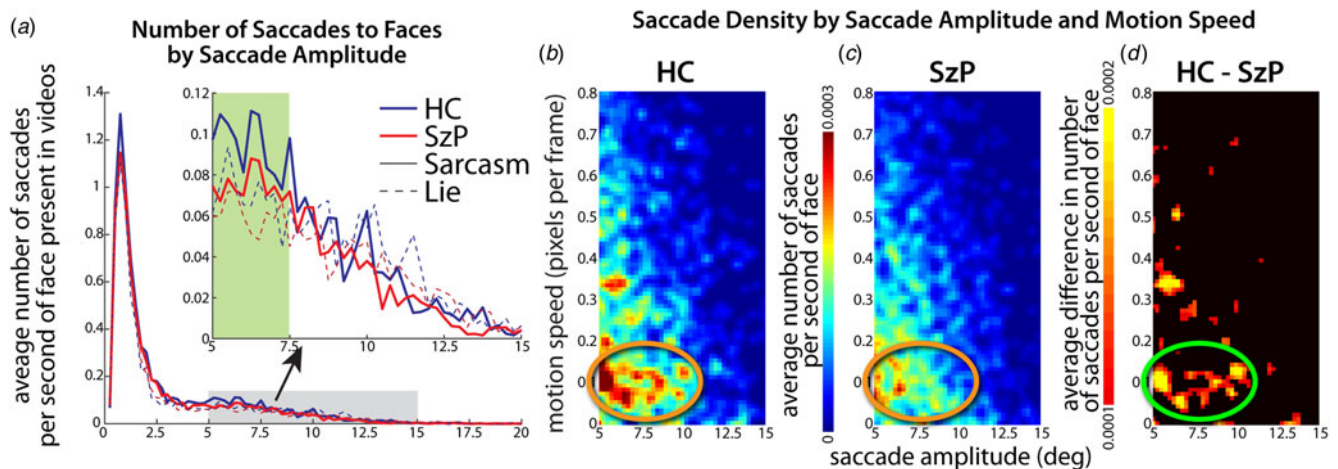


Fig. 4. Saccades to faces in peripheral vision. (a) Average number of saccades to faces by saccade amplitude for sarcasm and lie trials. Amplitude binned by 0.25°. Gray shaded area shows peripheral field counts highlighted in inset. Green shaded area in inset highlights the saccade amplitudes over which the number of HC saccades was significantly greater than SzP. (b and c) 2D histogram showing the density of saccades to the periphery (gray zone in a) for each group as a function of saccade amplitude and motion speed. Saccade amplitude was binned by 0.25°, and motion speed was binned by 0.0075 (1/80th of the 99.9th percentile of the motion distribution). Orange circles highlight the cluster of missing peripheral saccades in SzP v. HC. (d) Group difference map. Green circle highlights the cluster of missing peripheral saccades in SzP v. HC.

While this study provides a framework for how sensory processing deficits can impact social cognition and ultimately social functioning, there remain many gaps to fill. First is to understand what factors account for the ~50% in unexplained variance in TASIT performance between groups: SzP may also have additional impairments in making inferences about mental states that are independent of the sensory deficits and cognitive deficits measured by the PSI (Green & Horan, 2010; Pinkham et al., 2015; Savla, Vella, Armstrong, Penn, & Twamley, 2013).

Another potential source of intergroup variance is the impact of the chronicity of the disease. While medications do not appear to affect social functioning (Velthorst et al., 2017), many years of the lack of experience with social interactions may have exaggerated intergroup differences through decreased or altered use of the underlying brain circuits. Another gap is understanding what SzP are looking at *instead* of faces. This will require increased use of automated visual feature detection algorithms to further classify what objects and visual features are on screen at any given time. Another gap is in understanding how these findings, especially the biomarker-like separation of SzP and HC in Figs 1d and e, generalize and replicate, not only in a larger cohort but also with other naturalistic or real-world stimuli. In particular, longer naturalistic stimuli are needed: the short videos of the TASIT prevented us from assessing the number of saccades made to peripheral stimuli in individuals and directly measuring their impact on social cognition.

Lastly, these findings need to relate to the measures of brain pathology in SzP, including recent models of how excitatory-inhibitory circuit disturbances in SzP may underlie visual processing and cognitive operations (Anderson et al., 2016; Murray et al., 2014) and how brain areas involved in social cognition (such as those in the temporoparietal junction/posterior superior temporal sulcus, or TPJ-pSTS) may underlie visual scanning behavior (Corbetta et al., 2008; Green et al., 2015; Patel et al., 2019).

Although this study focused on SzP, similar approaches could also be applied to other groups with known social cognitive and functioning deficits (e.g. ASD) (Morrison et al., 2019; Veddem, Pedersen, Landert, & Bliksted, 2019). The use of naturalistic stimuli paired with the analytical techniques described here will increasingly serve as a bridge between the basic neuroscience literature and clinical studies of patient populations by providing smoothly varying behavioral measures that can be used as regressors to search for neural correlates (Jacoby, Bruneau, Koster-Hale, & Saxe, 2016; Russ & Leopold, 2015). These methods take advantage of not only increased computational power and the associated advances in computer vision, but also the vast literature in visual neuroscience collected over the past four decades. Moreover, the simplicity of administering tests based on naturalistic stimuli promises to produce low-burden and easily deployed clinical assessment tools that directly link the symptoms each individual is experiencing with underlying cognitive and neural

deficits, leading the way to individualized treatment regimens aimed at improving social functioning.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291720001646>

Acknowledgements. We wish to thank Rachel Marsh, Guillermo Horga, and Chad Sylvester for their comments on this manuscript.

Financial support. We wish to thank the funding agencies who supported this work: NIMH (GHP: K23MH108711 and T32MH018870; DCJ: R01MH049334; DAL and RAB: Intramural Research Program ZIA MH002898), Brain & Behavior Research Foundation (GHP), American Psychiatric Foundation (GHP), Sidney R. Baer Foundation (GHP), Leon Levy Foundation (GHP), and the Herb and Isabel Stusser Foundation (DCJ).

Conflict of interest. GHP receives income and equity from Pfizer, Inc through family. DCJ has equity interest in Glytech, AASI, and NeuroRx. He serves on the board of Promentis. He holds intellectual property rights for the use of NMDAR agonists in the treatment of schizophrenia, NMDAR antagonists in the treatment of depression and PTSD, and has submitted disclosures for fMRI-based prediction of ECT and TMS response, and EEG-based diagnosis of neuropsychiatric disorders. Within the past 2 years, he has received consulting payments/honoraria from Cadence, Biogen, SK Life Science, Autifony, Glytech and Boehringer Ingelheim. SCA, DRB, HMD, NES, LPB, JG, AM, RAB, and DAL reported no biomedical financial interests or potential conflicts of interest.

References

- Anderson, E. J., Tibber, M. S., Schwarzkopf, D. S., Shergill, S. S., Fernandez-Egea, E., Rees, G., & Dakin, S. C. (2016). Visual population receptive fields in people with schizophrenia have reduced inhibitory surrounds. *The Journal of Neuroscience*, *37*(6), 1546–1556. <http://doi.org/10.1523/JNEUROSCI.3620-15.2016>.
- Arnold, A. E. G. F., Iaria, G., & Goghari, V. M. (2016). Efficacy of identifying neural components in the face and emotion processing system in schizophrenia using a dynamic functional localizer. *Psychiatry Research: Neuroimaging*, *248*, 55–63. <http://doi.org/10.1016/j.psychres.2016.01.007>.
- Bulzacka, E., Meyers, J. E., Boyer, L., Le Gloahac, T., Fond, G., Szöke, A., ... Schürhoff, F. (2016). WAIS-IV seven-subtest short form: Validity and clinical use in schizophrenia. *Archives of Clinical Neuropsychology*, *31*(8), 915–925. <http://doi.org/10.1093/arclin/acw063>.
- Butler, P. D., Abeles, I. Y., Weiskopf, N. G., Tambini, A., Jalbrzikowski, M., Legatt, M. E., ... Javitt, D. C. (2009). Sensory contributions to impaired emotion processing in schizophrenia. *Schizophrenia Bulletin*, *35*(6), 1095–1107. <http://doi.org/10.1093/schbul/sbp109>.
- Butler, P. D., Zemon, V., Schechter, I., Saperstein, A. M., Hoptman, M. J., Lim, K. O., ... Javitt, D. C. (2005). Early-stage visual processing and cortical amplification deficits in schizophrenia. *Archives of General Psychiatry*, *62*(5), 495–504. <http://doi.org/10.1001/archpsyc.62.5.495>.
- Calderone, D. J., Martinez, A., Zemon, V., Hoptman, M. J., Hu, G., Watkins, J. E., ... Javitt, D. C. (2013). Comparison of psychophysical, electrophysiological, and fMRI assessment of visual contrast responses in patients with schizophrenia. *Neuroimage*, *67*, 153–162. <http://doi.org/10.1016/j.neuroimage.2012.11.019>.
- Chen, Y., Levy, D. L., Sheremata, S., & Holzman, P. S. (2004). Compromised late-stage motion processing in schizophrenia. *Biological Psychiatry*, *55*(8), 834–841. <http://doi.org/10.1016/j.biopsych.2003.12.024>.
- Chen, Y., Palafox, G. P., Nakayama, K., Levy, D. L., Matthyse, S., & Holzman, P. S. (1999). Motion perception in schizophrenia. *Archives of General Psychiatry*, *56*(2), 149–154.
- Corbetta, M., Patel, G. H., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*(3), 306–324. <http://doi.org/10.1016/j.neuron.2008.04.017>.
- Dias, E. C., Van Voorhis, A. C., Braga, F., Todd, J., Lopez-Calderon, J., Martinez, A., & Javitt, D. C. (2020). Impaired fixation-related theta modulation predicts reduced visual span and guided search deficits in schizophrenia. *Cerebral Cortex (New York, NY: 1991)*, *30*(5), 2823–2833. <http://doi.org/10.1093/cercor/bhz277>.
- Green, M. F., & Horan, W. P. (2010). Social cognition in schizophrenia. *Current Directions in Psychological Science*, *19*(4), 243–248. <http://doi.org/10.1177/0963721410377600>.
- Green, M. F., Horan, W. P., & Lee, J. (2019). Nonsocial and social cognition in schizophrenia: Current evidence and future directions. *World Psychiatry*, *18*(2), 146–161. <http://doi.org/10.1002/wps.20624>.
- Green, M. F., Horan, W. P., & Lee, J. (2015). Social cognition in schizophrenia. *Nature Reviews Neuroscience*, *16*(10), 620–631. <http://doi.org/10.1038/nrn4005>.
- Heimberg, C., Gur, R. E., Erwin, R. J., Shtasel, D. L., & Gur, R. C. (1992). Facial emotion discrimination: III. Behavioral findings in schizophrenia. *Psychiatry Research*, *42*(3), 253–265.
- Holdnack, J. A., Prifitera, A., Weiss, L. G., & Saklofske, D. H. (2015). Chapter 12. WISC-V and the personalized assessment approach. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V Assessment and Interpretation* (pp. 373–414). New York: Elsevier Inc.
- Jacoby, N., Bruneau, E., Koster-Hale, J., & Saxe, R. (2016). Localizing pain matrix and theory of mind networks with both verbal and non-verbal stimuli. *Neuroimage*, *126*, 39–48. <http://doi.org/10.1016/j.neuroimage.2015.11.025>.
- Javitt, D. C. (2009). When doors of perception close: Bottom-up models of disrupted cognition in schizophrenia. *Annual Review of Clinical Psychology*, *5*, 249–275. <http://doi.org/10.1146/annurev.clinpsy.032408.153502>.
- Johnston, P. J., Enticott, P. G., Mayes, A. K., Hoy, K. E., Herring, S. E., & Fitzgerald, P. B. (2010). Symptom correlates of static and dynamic facial affect processing in schizophrenia: Evidence of a double dissociation? *Schizophrenia Bulletin*, *36*(4), 680–687. <http://doi.org/10.1093/schbul/sbn136>.
- Kohler, C. G., Martin, E. A., Milonova, M., Wang, P., Verma, R., Bressinger, C. M., ... Gur, R. C. (2008). Dynamic evoked facial expressions of emotions in schizophrenia. *Schizophrenia Research*, *105*(1–3), 30–39. <http://doi.org/10.1016/j.schres.2008.05.030>.
- Kusunoki, M., Gottlieb, J. P., & Goldberg, M. E. (2000). The lateral intraparietal area as a salience map: The representation of abrupt onset, stimulus motion, and task relevance. *Vision Research*, *40*(10–12), 1459–1468. [http://doi.org/10.1016/S0042-6989\(99\)00212-6](http://doi.org/10.1016/S0042-6989(99)00212-6).
- Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughhead, J., Valdez, J. N., ... Gur, R. C. (2010). 'It's Not What You Say, But How You Say it': A reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, *4*, 19. <http://doi.org/10.3389/fnhum.2010.00019>.
- Leitman, D. I., Ziwich, R., Pasternak, R., & Javitt, D. C. (2006). Theory of Mind (ToM) and counterfactuality deficits in schizophrenia: Misperception or misinterpretation? *Psychological Medicine*, *36*(8), 1075–1083. <http://doi.org/10.1017/S0033291706007653>.
- Marsman, J.-B. C., Cornelissen, F. W., Dorr, M., Vig, E., Barth, E., & Renken, R. J. (2016). A novel measure to determine viewing priority and its neural correlates in the human brain. *Journal of Vision*, *16*(6), 3–3. <http://doi.org/10.1167/16.6.3>.
- Martinez, A., Gaspar, P. A., Hillyard, S. A., Andersen, S. K., Lopez-Calderon, J., Corcoran, C. M., & Javitt, D. C. (2018). Impaired motion processing in schizophrenia and the attenuated psychosis syndrome: Etiological and clinical implications. *American Journal of Psychiatry*, *175*(12), 1243–1254. <http://doi.org/10.1176/appi.ajp.2018.18010072>.
- McDonald, S., Bornhofen, C., Shum, D., Long, E., Saunders, C., & Neulinger, K. (2006). Reliability and validity of The Awareness of Social Inference Test (TASIT): A clinical test of social perception. *Disability and Rehabilitation*, *28*(24), 1529–1542. <http://doi.org/10.1080/09638280600646185>.
- McDonald, S., Flanagan, S., Rollins, J., & Kinch, J. (2003). TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, *18*(3), 219–238.
- Morrison, K. E., Pinkham, A. E., Kelsven, S., Ludwig, K., Penn, D. L., & Sasson, N. J. (2019). Psychometric evaluation of social cognitive measures for adults with autism. *Autism Research*, *12*(5), 766–778. <http://doi.org/10.1002/aur.2084>.
- Murray, J. D., Anticevic, A., Gancsos, M., Ichinose, M., Corlett, P. R., Krystal, J. H., & Wang, X.-J. (2014). Linking microcircuit dysfunction to cognitive

- impairment: Effects of disinhibition associated with schizophrenia in a cortical working memory model. *Cerebral Cortex (New York, NY: 1991)*, 24(4), 859–872. <http://doi.org/10.1093/cercor/bhs370>.
- Nieuwenhuys, R., Voogd, J., & van Huijzen, C. (2008). *The human central nervous system* (4 ed.). New York: Springer-Verlag.
- Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., ... Marder, S. R. (2008). The MATRICS Consensus Cognitive Battery, part 1: Test selection, reliability, and validity. *The American Journal of Psychiatry*, 165(2), 203–213. <http://doi.org/10.1176/appi.ajp.2007.07010042>.
- Okruszek, Ł., & Pilecka, I. (2017). Biological motion processing in schizophrenia – Systematic review and meta-analysis. *Schizophrenia Research*, 190, 3–10. <http://doi.org/10.1016/j.schres.2017.03.013>.
- Orbello, D. M., Grim, M. A., Talbott, R. E., & Ross, E. D. (2005). Impaired comprehension of affective prosody in elderly subjects is not predicted by age-related hearing loss or age-related cognitive decline. *Journal of Geriatric Psychiatry and Neurology*, 18(1), 25–32. <http://doi.org/10.1177/0891988704272214>.
- Patel, G. H., Sestieri, C., & Corbetta, M. (2019). The evolution of the temporoparietal junction and posterior superior temporal sulcus. *Cortex*, 118, 38–50. <http://doi.org/10.1016/j.cortex.2019.01.026>.
- Pinkham, A. E., Penn, D. L., Green, M. F., & Harvey, P. D. (2015). Social cognition psychometric evaluation: Results of the initial psychometric study. *Schizophrenia Bulletin*, 42(2), 494–504.
- Russ, B. E., & Leopold, D. A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *Neuroimage*, 109, 84–94. <http://doi.org/10.1016/j.neuroimage.2015.01.012>.
- Savla, G. N., Vella, L., Armstrong, C. C., Penn, D. L., & Twamley, E. W. (2013). Deficits in domains of social cognition in schizophrenia: A meta-analysis of the empirical evidence. *Schizophrenia Bulletin*, 39(5), 979–992. <http://doi.org/10.1093/schbul/sbs080>.
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., ... Tootell, R. B. H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science (New York, NY)*, 268(5212), 889–893.
- Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Current Biology: CB*, 20(7), 649–656. <http://doi.org/10.1016/j.cub.2010.02.032>.
- Sowden, S., Schuster, B., & Cook, J. (2019). The role of movement kinematics in facial emotion expression. Presented at the 11th International Conference on Educational Neuroscience. http://doi.org/10.3389/conf.fnhum.2019.229.00018/event_abstract.
- Sparks, A., McDonald, S., Lino, B., O'Donnell, M., & Green, M. J. (2010). Social cognition, empathy and functional outcome in schizophrenia. *Schizophrenia Research*, 122(1–3), 172–178. <http://doi.org/10.1016/j.schres.2010.06.011>.
- Taylor, S. F., Kang, J., Brege, I. S., Tso, I. F., Hosanagar, A., & Johnson, T. D. (2012). Meta-analysis of functional neuroimaging studies of emotion perception and experience in schizophrenia. *Biological Psychiatry*, 71(2), 136–145. <http://doi.org/10.1016/j.biopsych.2011.09.007>.
- Veddum, L., Pedersen, H. L., Landert, A.-S. L., & Bliksted, V. (2019). Do patients with high-functioning autism have similar social cognitive deficits as patients with a chronic cause of schizophrenia? *Nordic Journal of Psychiatry*, 73(1), 44–50. <http://doi.org/10.1080/08039488.2018.1554697>.
- Velthorst, E., Fett, A.-K. J., Reichenberg, A., Perlman, G., van Os, J., Bromet, E. J., & Kotov, R. (2017). The 20-year longitudinal trajectories of social functioning in individuals with psychotic disorders. *American Journal of Psychiatry*, 174(11), 1075–1085. <http://doi.org/10.1176/appi.ajp.2016.15111419>.
- Weschler, D. (1997). *WAIS-III Wechsler Adult Intelligence Scale*. San Antonio, TX: The Psychological Corporation.
- White, B. J., Berg, D. J., Kan, J. Y., Marino, R. A., Itti, L., & Munoz, D. P. (2017). Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature Communications*, 8(1), 14263. <http://doi.org/10.1038/ncomms14263>.
- Wilmington, N., Kietzmann, T. C., Jutras, M., Xue, C., Treue, S., Buffalo, E. A., & König, P. (2017). Differential contribution of low- and high-level image content to eye movements in monkeys and humans. <http://doi.org/10.1093/cercor/bhw399>.
- Zaki, J., & Ochsner, K. N. (2009). The need for a cognitive neuroscience of naturalistic social cognition. *Annals of the New York Academy of Sciences*, 1167, 16–30. <http://doi.org/10.1111/j.1749-6632.2009.04601.x>.
- Zhu, X., & Ramanan, D. (2012). *Face detection, pose estimation, and landmark localization in the wild*. Rhode Island: Computer Vision and Pattern Recognition (CVPR) Providence.