Supplementary Materials for the paper, *Spared bottom-up but impaired top-down interactive effects on naturalistic spoken language comprehension in schizophrenia.*

Hugh Rabagliati[1,2,3], Nathaniel Delaney-Busch[1], Jesse Snedeker[2] and Gina Kuperberg[1,4]

[1] Department of Psychology, Tufts University, Medford, MA 02155, U.S.A.

[2] Department of Psychology, Harvard University

[3] School of Philosophy, Psychology and Language Sciences, University of Edinburgh

[4] Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, U.S.A.

Table of Contents.

1. Full description of coding methods for eye movements: Video-based eye tracking

Although visual world eye tracking is new to the study of schizophrenia, it has considerable history in the field of psycholinguistics [1-9]. The variant used here – *video-based eye tracking,* in which gaze is recorded by a video camera – also has considerably history in developmental psychology, due to its non-invasive nature and its reliability (it results in minimal missing data). Video-based eye tracking has been used to study visual attention in both children and adults (e.g., [10]) but has mainly been used to study language. Important examples include a number of papers by Anne Fernald's research group (e.g., [11-13]). In particular, this research group has demonstrated that video-based eye tracking is a valid measure of language comprehension, by showing that children's eye-tracked responses to spoken sentences can predict their scores on standardized structural language assessments completed at both the time of testing and also many years later [14].

Other groups have also used this method (e.g., [15, 16]). The results from [16] are particularly important here because they conducted a comparison between automated eye tracking and video-based eye tracking, and found that the two techniques measured position of gaze with greater than 90% agreement. As such, video-based eye tracking can be described as a robust and non-invasive approach to measure gaze.

In the current study, eye movement videos were coded to reveal how participants gazed at the different objects in front of them over time, and how this varied across the experimental conditions. All coding was carried out by trained coders employed by the lab; these coders were blind to hypotheses and experimental conditions and were not otherwise involved in the running of this study. All coders had taken part in a rigorous training routine, coding multiple different videos iteratively in order to achieve high across-coder reliability (>95% agreement across five different videos).

Eye movement coding was carried out, frame-by-frame, starting from the onset of the first word in each sentence. For instance, for the trial associated with the spoken instruction, "*tickle the frog with the feather*", one coder would first find the onset of the word "*tickle*" in the video, and then (with the sound turned off), a second coder (blind to condition) would record the participant's direction of gaze in each video frame from that point up until 2500ms subsequent to the onset of the final word ("*feather*"). Although this procedure recorded information for the entire duration of the sentence, we only analyzed gaze on frames subsequent to the onset of "*feather*", as that was the point at which participants could begin to resolve the syntactic ambiguity to come to interpretation of sentence meaning.

To code each frame of the video, the coder recorded where the participant was gazing on the shelf (to the upper left object, upper right object, lower left object, lower right object, the center of the shelf, elsewhere on the shelf), or recorded if the subject blinked or looked away from the shelf entirely. Once each frame had been coded, the subsequent file was recoded to match the location of gaze with the object of gaze (e.g., on trials where the upper left object was the Target Instrument, frames with gaze to the upper left were recoded as gaze to the Target Instrument). Because our critical hypotheses concerned gaze to the Target Instrument (as a measure of

syntactic ambiguity resolution), our key analyses compared gaze to that object across the different experimental conditions.

The resulting data were then analyzed using the "time-window" analysis reported in the main paper. This analysis compared gaze to the Target Instrument across conditions in two pre-specified a priori 500ms time windows, from 200ms-699ms after the onset of the final word (e.g., "*feather*"), and from 700ms-1199ms. For each trial and time window, we coded whether or not participants fixated on the Target Instrument, to create the binary dependent measure that could be analyzed using the logistic regressions described in the main paper (using the formulae alongside the Tables starting on p.9).

2. Graphs depicting time-course of gaze across the different objects in Tasks 1-3
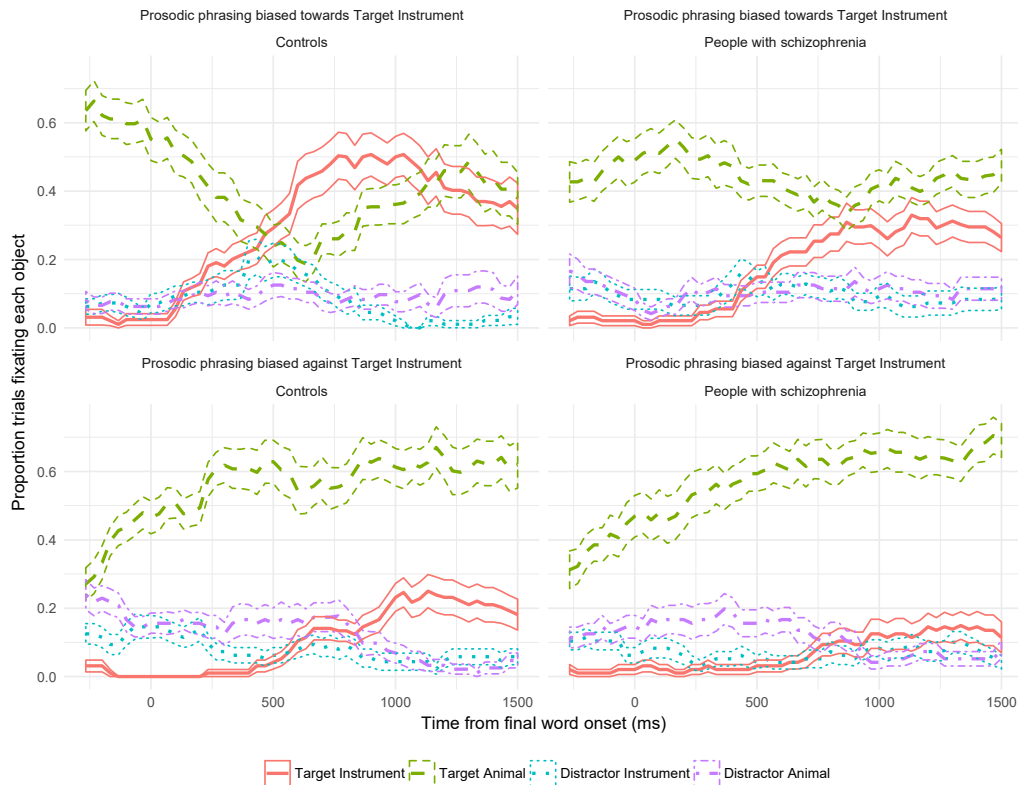
Supplementary Figure 1. Task 1: Use of prosodic phrasing over time in people with schizophrenia and controls. For each condition, the graphs show proportion of trials, at each time point, on which participants are fixating upon each of the objects. Ribbons represent +/- one standard error of the mean. Note that target and distractor animals were of different types (e.g., a frog and a cat).

An example stimulus set would be:

Biased towards target instrument: *You can point at the cat... with the flower*
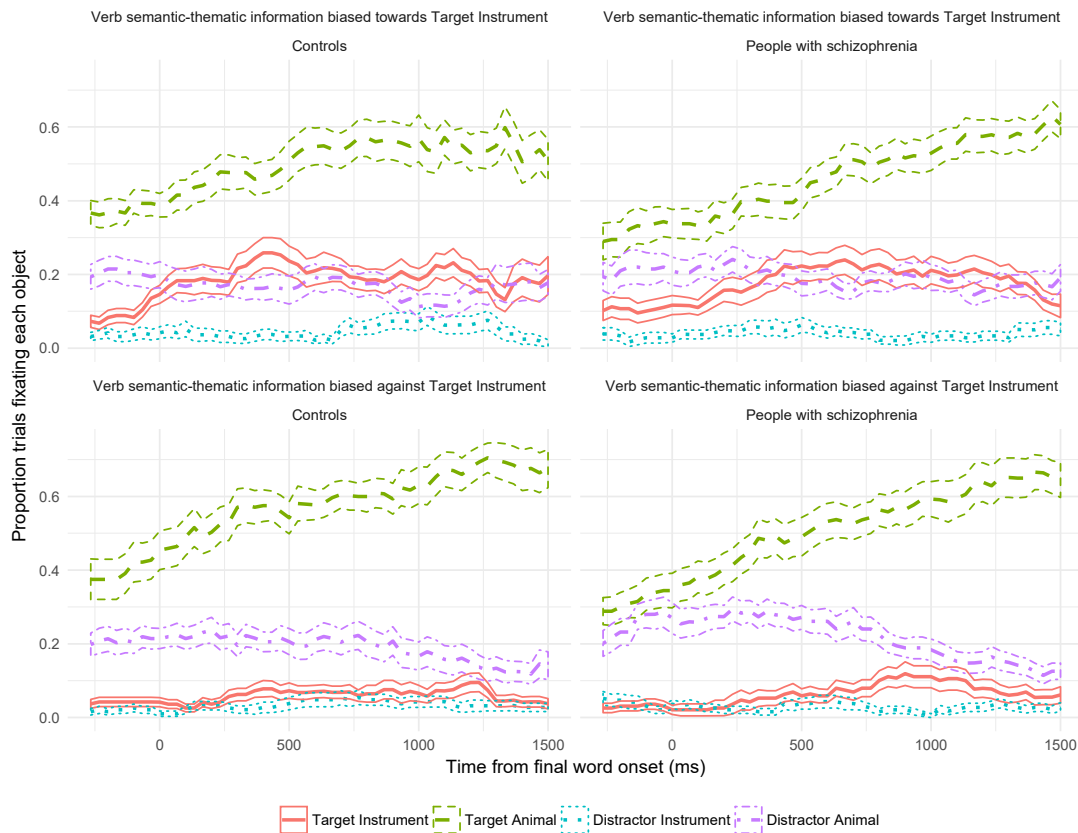
Biased against target instrument: *You can point at... the cat with the flower*

Target Animal: A cat holding a tiny flower

Target Instrument: A large flower

Distractor Animal: An elephant holding a tiny hairclip

Distractor Instrument: A large hairclip

Supplementary Figure 2. Task 2 (Verb semantic-thematic): Use of verb semantic-thematic information over time in people with schizophrenia and controls. For each condition, the graphs show proportion of trials, at each time point, on which participants are fixating each of the objects. Ribbons represent +/- one standard error of the mean. Target and distractor animals were of the same type on 50% of trials.

An example stimulus set would be (for trials with two different type of animal):

Biased towards target instrument:

Poke the penguin with the feather

Target Animal: A penguin holding a tiny feather

Target Instrument: A large feather

Distractor Animal: A monkey holding a tiny partyhat

Distractor Instrument: A large partyhat

Biased against target instrument:

Sing to the koala with the funnel

Target Animal: A koala holding a tiny funnel

Target Instrument: A large funnel

Distractor Animal: A dinosaur holding a tiny flower

Distractor Instrument: A large flower

Supplementary Figure 3. Task 2 (Visual Scene): Use of pragmatically-relevant visual scene information over time in people with schizophrenia and controls. For each condition, the graphs show proportion of trials, at each time point, on which participants are fixating each of the objects. Ribbons represent +/- one standard error of the mean. Target and distractor animals were of different types when the scene was biased toward using the target instrument, and of the same type when biased against using the target instrument.

An example stimulus set would be:

Biased towards target instrument:

Poke the penguin with the feather

Target Animal: A penguin holding a tiny feather

Target Instrument: A large feather

Distractor Animal: A monkey holding a tiny partyhat

Distractor Instrument: A large partyhat
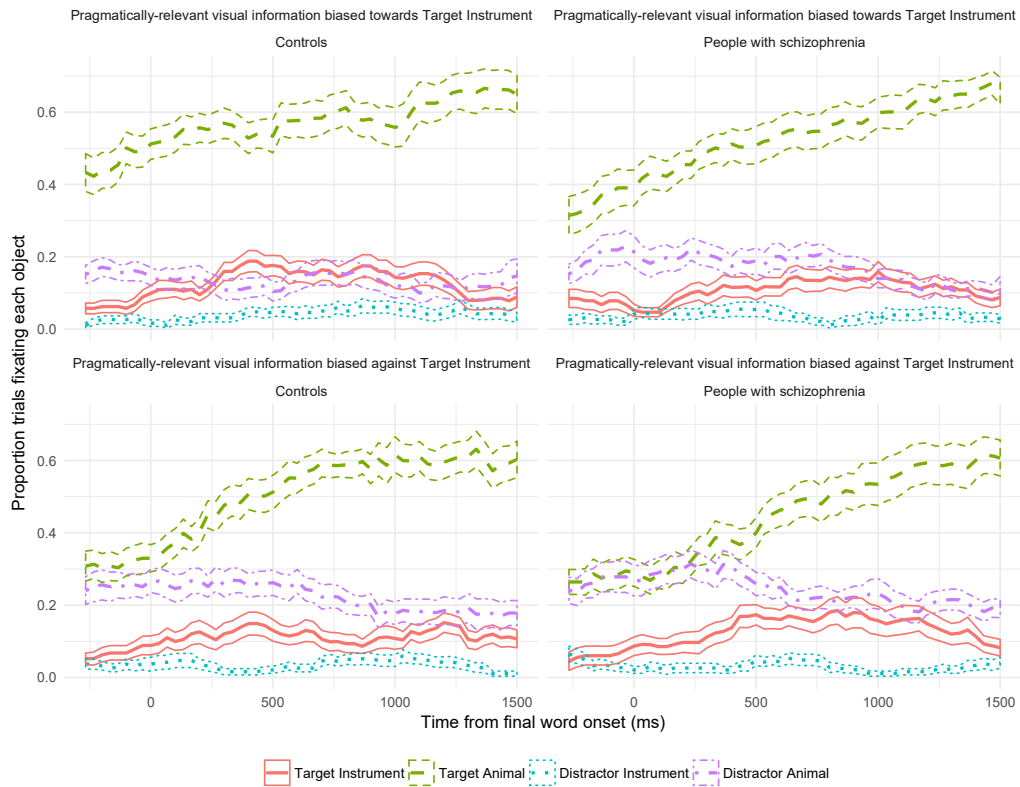
Biased against target instrument:
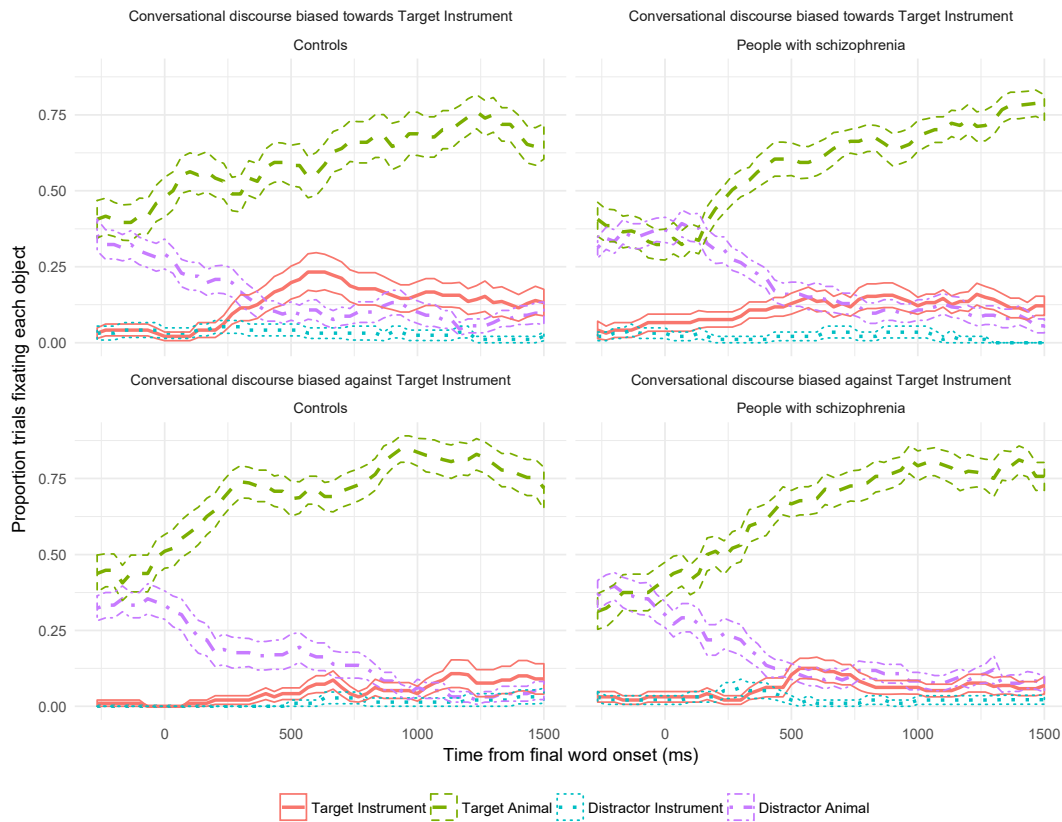
Poke the penguin with the feather

Target Animal: A penguin holding a tiny feather

Target Instrument: A large feather

Distractor Animal: A penguin holding a tiny partyhat

Distractor Instrument: A large partyhat

Supplementary Figure 4. Task 3: Use of conversational discourse information over time in people with schizophrenia and controls. For each condition, the graphs show proportion of trials, at each time point, on which participants are fixating each of the objects. Ribbons represent +/- one standard error of the mean. Note that target and distractor animals were always of the same type.

An example stimulus set would be:

Biased towards target instrument: Male Speaker: *What should we do to a fox?*

Female Speaker: *Scratch the fox with the comb.*

Biased against target instrument: Male Speaker: *Which fox should we use now?*

Female Speaker: *Scratch the fox with the comb.*

Target Animal: A fox holding a tiny comb

Target Instrument: A large comb

Distractor Animal: A fox holding a tiny ball

Distractor Instrument: A large ball

3. Full inferential statistics for the logistic regression analyses of eye movements.

Table S1. Full set of inferential statistics from the omnibus logistic regression analyses of participants' eye movements. All models were calculated using lme4 package version 1.1 [17] in R [18].

Task 1's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Group * Prosodic Phrasing * Time Window + (1+ Time Window |Participant)*

Task 2's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Group * Verb Information * Visual Information * Time Window + (1+ Verb Information + Visual Information + Time Window |Participant)*

Task 3's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Group * Conversational Discourse * Time Window + (1+ Time Window |Participant)*

Task 1 (Manipulation of prosodic phrasing)	β Estimate	Std. Error	Wald *z* statistic	*p* value (* <.05)
(Intercept)	-0.60	0.20	-3.00	.003
Effect of Group	-1.04	0.30	-3.51	<.001 *
Effect of Prosodic Phrasing	-0.82	0.13	-6.41	<.001 *
Effect of Time Window	0.52	0.13	4.04	<.001 *
Group * Prosodic Phrasing	0.11	0.20	0.53	.593
Group * Time Window	0.05	0.20	0.23	.820
Prosodic Phrasing * Time Window	0.09	0.13	0.73	.466
Group * Prosodic Phrasing * Time Window	0.01	0.20	0.03	.976

Task 2 (Manipulation of the verb's semantic-thematic information & pragmatically relevant visual information)

(Intercept)	-1.61	0.17	-9.64	< 2e-16
Effect of Group	0.03	0.15	0.22	.825
Effect of Verb Information	-0.88	0.12	-7.11	<.001 *
Effect of Visual Information	-0.14	0.10	-1.38	.167
Effect of Time Window	0.10	0.08	1.21	.225
Group * Verb Information	0.07	0.11	0.68	.499
Group * Visual Information	0.21	0.10	2.12	.034 *
Verb Information * Visual Information	-0.15	0.08	-1.82	.070
Group * Time Window	0.00	0.07	-0.05	.957
Verb Information * Time Window	0.11	0.08	1.38	.169
Visual Information * Time Window	-0.02	0.08	-0.24	.810
Group * Verb Information * Visual Information	0.05	0.08	0.61	.543
Group * Verb Information * Time Window	-0.01	0.07	-0.07	.943
Group * Visual Information * Time Window	-0.02	0.07	-0.27	.785
Verb Information * Visual Information * Time Window	0.04	0.07	0.48	.634
Group * Verb Information * Visual Information * Time Window	-0.08	0.07	-1.14	.254

Task 3 (Manipulation of conversational discourse)

(Intercept)	-1.72	0.23	-7.40	<.001 *
Effect of Group	0.02	0.31	0.07	.947
Effect of Conversational Discourse	-0.55	0.15	-3.72	<.001 *
Effect of Time Window	0.25	0.16	1.57	.116
Group * Conversational Discourse	0.44	0.20	2.19	.029 *
Group * Time Window	-0.09	0.20	-0.45	.655
Conversational Discourse * Time Window	0.19	0.15	1.27	.203
Group * Conversational Discourse * Time Window	-0.20	0.20	-0.98	.325

Table S2. Inferential statistics from the logistic regression analyses of Control

participants' eye movements.

Task 1's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Prosodic Phrasing * Time Window + (1+ Time Window*

|Participant)

Task 2's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Verb Information * Visual Information * Time Window + (1+*

Verb Information + Visual Information + Time Window |Participant)

Task 3's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Conversational Discourse * Time Window + (1+ Time Window*

|Participant)

Task 1 (Manipulation of prosodic phrasing)	β Estimate	Std. Error	Wald *z* statistic	*p* value (* <.05)
(Intercept)	-0.58	0.17	-3.40	.001 *
Effect of Prosodic Phrasing	-0.80	0.13	-6.27	<.001 *
Effect of Time Window	0.51	0.13	3.94	<.001 *
Prosodic Phrasing * Time Window	0.09	0.13	0.73	.468

Task 2 (Manipulation of the verb's semantic-thematic information & pragmatically relevant visual information)				
(Intercept)	-1.61	0.23	-6.96	<.001 *
Effect of Verb Information	-0.92	0.16	-5.74	<.001 *
Effect of Visual Information	-0.39	0.16	-2.37	.018 *
Effect of Time Window	0.15	0.12	1.21	.228
Verb Information * Visual Information	-0.22	0.12	-1.82	.069
Verb Information * Time Window	0.14	0.11	1.23	.219
Visual Information * Time Window	0.02	0.11	0.18	.857

Verb Information * Visual Information * Time Window	0.13	0.11	1.20	.230
Task 3 (Manipulation of conversational discourse information)				
(Intercept)	-1.86	0.30	-6.19	<.001 *
Effect of Conversational Discourse	-0.58	0.15	-3.82	<.001 *
Effect of Time Window	0.33	0.19	1.72	.085
Conversational Discourse * Time Window	0.21	0.15	1.38	.169

Table S3. Inferential statistics from the logistic regression analyses of Patient participants' eye movements.

Task 1's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Prosodic Phrasing * Time Window + (1+ Time Window*

|Participant)

Task 2's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Verb Information * Visual Information * Time Window + (1+*

Verb Information + Visual Information + Time Window |Participant)

Task 3's regression had the form (in lmer syntax)

*Gaze to Instrument ~ Conversational Discourse * Time Window + (1+ Time Window*

|Participant)

Task 1 (Manipulation of prosodic phrasing)	β Estimate	Std. Error	Wald *z* statistic	*p* value (* <.05)
(Intercept)	-1.72	0.27	-6.36	<.001 *
Effect of Prosodic Phrasing	-0.74	0.16	-4.73	<.001 *
Effect of Time Window	0.60	0.18	3.27	.001 *
Prosodic Phrasing * Time Window	0.10	0.16	0.65	.517

Task 2 (Manipulation of the verb's semantic-thematic information & pragmatically relevant visual information)				
(Intercept)	-1.61	0.24	-6.69	<.001 *
Effect of Verb Information	-0.84	0.19	-4.48	<.001 *
Effect of Visual Information	0.10	0.13	0.72	.474
Effect of Time Window	0.06	0.12	0.52	.601
Verb Information * Visual Information	-0.08	0.11	-0.72	.471
Verb Information * Time Window	0.09	0.11	0.86	.392
Visual Information * Time Window	-0.04	0.10	-0.42	.672
Verb Information * Visual	-0.05	0.10	-0.53	.598

Information * Time Window				

Task 3 (Manipulation of conversational discourse)				
(Intercept)	-1.62	0.18	-9.11	<.001 *
Effect of Conversational Discourse	-0.10	0.14	-0.76	.448
Effect of Time Window	0.12	0.15	0.85	.393
Conversational Discourse * Time Window	-0.01	0.14	-0.09	.927

4. Full inferential statistics for the logistic regression analyses of actions.

Table S4. Full set of inferential statistics from the omnibus logistic regression analyses of participants' actions.

Task 1's regression had the form (in lmer syntax)

*Use of Instrument ~ Group * Prosodic Phrasing + (1 |Participant)*

Task 2's regression had the form (in lmer syntax)

*Use of Instrument ~ Group * Verb Information * Visual Information + (1+ Verb Information + Visual Information |Participant)*

Task 3's regression had the form (in lmer syntax)

*Use of Instrument ~ Group * Conversational Discourse + (1 |Participant)*

Task 1 (Manipulation of prosodic phrasing)	β Estimate	Std. Error	Wald z statistic	p value (* <.05)
(Intercept)	-0.09	0.22	-0.40	.686
Effect of Group	-0.29	0.22	-1.35	.179
Effect of Prosodic Phrasing	-0.99	0.14	-7.17	<.001 *
Group * Prosodic Phrasing	-0.04	0.13	-0.32	.748

Task 2 (Manipulation of the verb's semantic-thematic information & pragmatically relevant visual information)				
(Intercept)	-1.54	0.29	-5.39	<.001 *
Effect of Group	-0.20	0.24	-0.80	.423
Effect of Verb Information	-1.60	0.22	-7.18	<.001 *
Effect of Visual Information	-0.19	0.14	-1.44	.151
Group * Verb Information	-0.19	0.17	-1.15	.252
Group * Visual Information	0.04	0.12	0.32	.750
Verb Information * Visual Information	0.06	0.13	0.42	.675

Group * Verb Information * Visual Information	0.12	0.11	1.06	.290
Task 3 (Manipulation of conversational discourse)				
(Intercept)	-1.54	0.29	-5.39	<.001 *
Effect of Group	-0.20	0.24	-0.80	.423
Effect of Conversational Discourse	-1.60	0.22	-7.18	<.001 *
Group * Conversational Discourse	-0.19	0.14	-1.44	.151

Table S4. Inferential statistics from the logistic regression analyses of Control group's actions.

Task 1's regression had the form (in lmer syntax)

Use of Instrument ~ Prosodic Phrasing + (1 |Participant)

Task 2's regression had the form (in lmer syntax)

*Use of Instrument ~ Verb Information * Visual Information + (1+ Verb Information + Visual Information |Participant)*

Task 3's regression had the form (in lmer syntax)

Use of Instrument ~ Conversational Discourse + (1 |Participant)

Task 1 (Manipulation of prosodic phrasing)	β Estimate	Std. Error	Wald *z* statistic	*p* value (* <.05)
(Intercept)	-0.35	0.23	-1.56	.118
Effect of Prosodic Phrasing	-0.94	0.18	-5.23	<.001 *
Task 2 (Manipulation of the verb's semantic-thematic information & pragmatically relevant visual information)				
(Intercept)	-1.13	0.26	-4.42	<.001 *
Effect of Verb Information	-1.19	0.20	-6.03	<.001 *
Effect of Visual Information	-0.24	0.17	-1.41	.159
Verb Information * Visual Information	-0.07	0.16	-0.43	.664
Task 3 (Manipulation of conversational discourse)				
(Intercept)	-1.86	0.30	-6.19	<.001 *
Effect of Conversational Discourse	-0.58	0.15	-3.82	<.001 *

Table S5. Inferential statistics from the logistic regression analyses of the

Schizophrenia group's actions.

Task 1's regression had the form (in lmer syntax)

Use of Instrument ~ Prosodic Phrasing + (1 |Participant)

Task 2's regression had the form (in lmer syntax)

*Use of Instrument ~ Verb Information * Visual Information + (1+ Verb Information +*

Visual Information |Participant)

Task 3's regression had the form (in lmer syntax)

Use of Instrument ~ Conversational Discourse + (1 |Participant)

Task 1 (Manipulation of prosodic phrasing)	β Estimate	Std. Error	Wald *z* statistic	*p* value (* <.05)
(Intercept)	0.25	0.40	0.61	.541
Effect of Prosodic Phrasing	-1.06	0.21	-4.95	<.001 *
Task 2 (Manipulation of the verb's semantic-thematic information & pragmatically relevant visual information)				
(Intercept)	-2.44	0.78	-3.14	.002 *
Effect of Verb Information	-2.50	0.67	-3.74	<.001 *
Effect of Visual Information	-0.10	0.22	-0.45	.653
Verb Information * Visual Information	0.24	0.22	1.08	.280
Task 3 (Manipulation of conversational discourse)				
(Intercept)	-1.60	0.32	-5.05	<.001 *
Effect of Conversational Discourse	-0.58	0.20	-2.87	.004 *

5. Exploratory correlations with clinical measures within the patient group

Within the schizophrenia group, we examined relationships between participants' use of different informational cues and their clinical characteristics. Given the number of participants and relatively small range of symptoms, we consider these analyses exploratory. We hypothesized that negative thought disorder severity (SANS global alogia rating) might predict impairments using prosodic phrasing. This was based on previous reports of correlations between negative symptoms and lower-level perceptual deficits [19] or prosodic impairments [20]. In addition, we hypothesized that positive thought disorder severity (SAPS global thought disorder rating) might predict impairments in using lexical information. This was based on previous reports of lexico-semantic abnormalities in individuals with positive thought disorder [e.g. 21, 22]. In addition, to establish the specificity of any findings, we carried out analyses examining global ratings of delusions, hallucinations, and global summed ratings of positive and negative symptoms, as well as with medication dosage.

We found that individuals with higher ratings of negative thought disorder (M=0.8, range 0 - 4) showed marginally less influence of prosodic information when carrying out their final actions (Beta=0.39 (Standard Error = 0.21), z=1.9, p=.06), although this effect was not found in eye movements (p=.52).

We also found that individuals with higher global positive thought disorder ratings (M=4.4, range 0-12) showed weaker effects of lexical information on both their eye movements (Beta=0.32 (0.15), z=2.1, p=.04) and final actions (Beta=0.74 (0.35), z=2.1, p=.03). While these findings should be considered preliminary, they are consistent with the theory that lexical representations are noisier in people with positive thought disorder as a longer-term consequence of compromised top-down interactions [23].

Global ratings of delusions, hallucinations, or summed scores of global ratings of positive and negative symptoms all failed to predict how any type of information

affected either eye movements or final actions. The same was true of medication

(chlorpromazine equivalents).

References

1.   Allopenna, P.D., J.S. Magnuson, and M.K. Tanenhaus, *Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models.* Journal of Memory and Language, 1998. **38**: p. 419-439.
2.   Dahan, D. and M.K. Tanenhaus, *Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2004. **30**(2): p. 498.
3.   McMurray, B., et al., *Gradient sensitivity to within-category variation in words and syllables.* J Exp Psychol Hum Percept Perform, 2008. **34**(6): p. 1609-31.
4.   Sedivy, J.C., et al., *Achieving incremental semantic interpretation through contextual representation.* Cognition, 1999. **71**(2): p. 109-147.
5.   Spivey, M.J., et al., *Eye movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution.* Cogn Psychol, 2002. **45**(4): p. 447-81.
6.   Tanenhaus, M.K., et al., *Integration of visual and linguistic information in spoken language comprehension.* Science, 1995. **268**(5217): p. 1632-4.
7.   Tanenhaus, M.K. and J.C. Trueswell, *Eye movements and spoken language comprehension.*, in *Handbook of Psycholinguistics*, M. Traxler and M. Gernsbacher, Editors. 2006, OUP: Oxford. p. 863-900.
8.   Altmann, G. and Y. Kamide, *Incremental interpretation at verbs: Restricting the domain of subsequent reference.* Cognition, 1999. **73**(3): p. 247-264.
9.   Huettig, F., J. Rommers, and A.S. Meyer, *Using the visual world paradigm to study language processing: A review and critical evaluation.* Acta psychologica, 2011. **137**(2): p. 151-171.
10.   Hood, B.M. and J. Atkinson, *Disengaging visual attention in the infant and adult.* Infant Behavior and Development, 1993. **16**(4): p. 405-422.
11.   Fernald, A., et al., *Rapid gains in speed of verbal processing by infants in the 2nd year.* Psychological Science, 1998: p. 228-231.
12.   Swingley, D., J.P. Pinto, and A. Fernald, *Continuous processing in word recognition at 24 months.* Cognition, 1999. **71**(2): p. 73-108.
13.   Fernald, A., D. Swingley, and J.P. Pinto, *When half a word is enough: Infants can recognize spoken words using partial phonetic information.* Child development, 2001. **72**(4): p. 1003-1015.
14.   Marchman, V.A. and A. Fernald, *Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood.* Developmental science, 2008. **11**(3): p. F9-F16.
15.   Song, H.-j. and C. Fisher, *Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns.* Journal of Memory and Language, 2005. **52**(1): p. 29-57.
16.   Snedeker, J. and J.C. Trueswell, *The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing.* Cognitive Psychology, 2004. **49**(3): p. 238-299.
17.   Bates, D.M., et al., *Fitting linear mixed-effects models using lme4.* Journal of Statistical Software, 2015. **67**(1): p. 1-48.

18.   R Core Team, *R: A language and environment for statistical computing*. 2015, ISBN 3-900051-07-0: Vienna, Austria.

19.   Leitman, D.I., et al., *Sensory contributions to impaired prosodic processing in schizophrenia.* Biological Psychiatry, 2005. **58**(1): p. 56-61.

20.   Jahshan, C., J.K. Wynn, and M.F. Green, *Relationship between auditory processing and affective prosody in schizophrenia.* Schizophrenia Research, 2013. **143**(2-3): p. 348-353.

21.   Spitzer, M., et al., *Associative semantic network dysfunction in thought-disordered schizophrenic patients: Direct evidence from indirect semantic priming.* Biological Psychiatry, 1993. **34**: p. 864-877.

22.   Kreher, D.A., D. Goff, and G.R. Kuperberg, *Why all the confusion? Experimental task explains discrepant semantic priming effects in schizophrenia under "automatic" conditions: Evidence from event-related potentials.* Schizophrenia Research, 2009. **111**(1-3): p. 174-181.

23.   Brown, M. and G.R. Kuperberg, *A hierarchical generative framework of language processing: Linking language perception, interpretation, and production abnormalities in schizophrenia.* Frontiers in Human Neuroscience, 2015. **9**: p. 643.