Supplemental Materials

Supplemental Methods - participant demographics

Full-scale and Verbal IQ scores were obtained from all participants using the Wechsler Abbreviated Scale of Intelligence. One participant's Verbal IQ score was not obtained because their IQ was tested and scored off-site, and the data were not available. Informed assent or consent was obtained from all participants (or their legal guardian when appropriate) in accordance with a National Institutes of Health Institutional Review Board approved protocol. Handedness was obtained for participants using the Physical and Neurological Examination for Soft Signs (Denckla, 1985) in which participants were asked which hand they used for 12 everyday objects. Scores of 10 or more qualified for right- or left-handedness and lesser scores qualified for ambidexterity. The autism and control groups had similar numbers of right and non-right handers: The control group had 17 right-handers and 3 left-handers. The autism group had 16 right-handers and 3 ambidextrous participants.

Behavioral data collection

The participant was equipped with an Opto-Acoustics FOMRI-III NC Microphone with built-in noise cancellation and over-the-ear headphones. The experimenter wore a lapel-clipped Shure MX-184 microphone and in-ear Etymotic Research headphones set. Audio signals from the microphones were routed with an M-Audio FastTrack Ultra 8-R USB sound card and saved to uncompressed WAV files with Adobe Audition. A video image of the experimenter was captured with a Panasonic HC-V720P digital high definition camera mounted directly in front of the experimenter at eye-level, approximately 1.5 meters away. The 800x600 pixel video stream of the experimenter was transmitted to an MRI-compatible NordicNeuroLab VisualSystem goggle display mounted directly to the head coil. This allowed the subject a full view of the experimenter during the Conversation and Repetition scans. A video image of subject's hands was captured by a camera mounted permanently within the MRI chamber, and both video streams were saved to a Samsung Digital Video Recorder. The images on the display of the laptop presenting instructions were also recorded.

Behavioral data processing

To ensure precise measurements of when speech occurred, we reduced the scanner background noise present in the recordings, and adjusted the dynamic range (loudness) of the speech using the "Noise-Removal", "Amplify-Compressor" and "Pop-Mute" tools from Audacity software. This was done on an *ad hoc* basis, but the general procedure was to first subtract the background noise using the standard “Noise-Removal” function, and selecting a portion of the WAV file that did not contain any speech to use as the background referent. Then we removed high amplitude “clicks” from the sound files using Pop-Mute (threshold = 6.00; look=10.00; rel = 10.00) then Amplify-Compressor was used to normalize the amplitude of the WAV file (threshold = 10.00; NoiseFloor = -40.00; Ratio = 9.5; AttackTime = 0.2; DecayTime = 1.00; Normalize = yes; UsePeak=no).

Number and duration of speaking turns were established using a Matlab script to detect times when the subject or experimenter was speaking, based on the amplitude of the cleaned WAV files. The script counted as “speech” any time point whose amplitude was greater than 10% of the maximum amplitude of the entire recording, and so long as its duration was at least 50 milliseconds. After this automatic process, all the resulting binary time series (1 for speech, 0 elsewhere) were plotted and visually inspected to ensure the automatic process had succeeded. Audio from the conversations was anonymized and transcribed by a commercial transcription service.

Effects of Listening and of Speaking

We used typical general linear model analyses with beta weights to examine differences in activity level during periods of speaking versus listening (Agnew *et al.*, 2013; Blank *et al.*, 2002; Price, 2012; Wise *et al.*, 1999). Because our conversations occurred spontaneously, events or blocks where speech occurred could not be defined in advance. Instead we defined speech perception and production blocks post hoc from the audio recordings of the conversations. First, the binary time series of speaking turn occurrences was down-sampled to seconds. For analysis, we selected all uninterrupted conversational speaking turns from both talkers that were at least 4 seconds in duration. These were assigned to two conditions: 1) "Speaking" (when participants spoke) and 2) "Listening" (when the conversation partner spoke). Segments shorter than 4 seconds and those that contained overlapping speech from the experimenter and subject were assigned to a third condition: 3) "Mixed", which was treated as a regressor of no interest. These 3 condition regressors were convolved with a hemodynamic response function using AFNI's Waver program and were analyzed with a general linear model using AFNI's 3dDeconvolve. The rest periods at the start and end of each run were not associated with a condition regressor and therefore were counted as part of the model baseline. The lack of rigorously spaced null or baseline periods during naturalistic conversation resulted in a baseline mean that included task-driven activity, allowing only the detection of relative differences between the conditions.

Individual subject betas were subject to a mixed effects model (3dLME) with Group (Autism/Control) and Turn (Speaking/Listening) as fixed factors and Subject as random intercept. We tested for main effects of Speaking > 0, Listening > 0, and the group effect of Control > Autism, and the interaction of Group X Turn at cluster corrected threshold of height $p<.001$ (FWE $p<.05$) (see Supplementary Fig. 3).

Supplementary Table 1: Participant demographics and psychometrics for Task Study

	ASD (n=19)	TD (n=20)	t(38)	p
WASI FSIQ (SS)[1]	112.8±12	118.7±8.9	1.7	0.09
WASI VIQ (SS)[2]	111.6±13.2	118.6±9.2	1.8	0.08
Age (years)	20±3.6	22±4.7	15	0.15
ADOS[3] (Soc+Comm raw score)	11.3±3.4			
SRS[4] Total raw score	98.7±28.2			

Supplementary Table 2: Counts of the topics of conversation, separated by diagnostic group. These apply to the first 2 topics of conversation only. The final topic of conversation always centered around work or school life. Two conversations in the control group have been excluded from this and all other analyses because a computer error caused the audio not to be saved.

Topic	Control	Autism
Art	1	0
Music	8	4
Games	4	7
TV/Movies	5	3
Cars	1	0
Computers/Internet	2	4
Literature, Books, Comics	3	6
Sports	1	0
Academic Subject (e.g. History, Math)	4	4
People/Social	4	3
Engineering/How things work	3	4

[1] Wexler's Abbreviated Scale of Intelligence, Full Scale IQ, Standard Score
[2] Wexler's Abbrevaited Scale of Intelligence, Verbal IQ, Standard Score
[3] Autism Diagnostic Observation Schedule
[4] Social Responsiveness Scale

Politics	1	0
Animals/Insects	1	3
Work/school life	20	19
TOTAL	**58**	**57**

Supplementary Table 3: Contrast of whole-brain correlation values, Conversation > Repetition

	Center of Mass Coordinates				
Location	**x**	**y**	**z**	**Peak Z**	**mm^3**
B/L dorsomedial PFC	2	29	45	6.232	7587
L frontotemporal	-35	17	-22	6.5985	7371
L angular gyrus	-47	-62	30	7.3455	6399
R premotor / IFG	47	5	33	6.8742	5670
R lenticular formation	23	2	-4	6.7702	5238
B/L dorsomedial PFC	-2	53	18	6.0746	2133
L posterior insula	-35	-14	12	6.0838	1485
L gingual lyrus	-14	-53	6	5.5424	1026
L middle temporal gyrus	-44	-29	-1	5.6953	999
L amygdala / subcallosal gyr	-17	-2	-19	5.6453	621
R anterior insula	38	17	-7	5.1811	594
R Heschls gyrus	35	-29	15	5.6532	567
L superior frontal gyrus gyrus	-20	17	48	5.4723	513
L posterior cingulate	-2	-41	30	5.6621	486
R angular gyrus	41	-56	33	5.2817	459
L putamen	-23	-2	6	5.253	405
R middle temporal gyrus	50	-29	-7	5.4095	324

Supplementary Table 4: Contrast of whole-brain correlation values, Autism>Control

	Peak coordinates				
Location	**x**	**y**	**z**	**Peak Z**	**mm^3**
R superior temporal sulcus	44	-20	-7	4.42	918
R superior temporal	53	8	-19	4.63	864

sulcus/MTG					
R ventral striatum	17	8	-13	4.36	783
R IFG/insula	44	5	9	4.82	702
L postcentral gyrus	-38	-29	51	3.85	675
L temporal pole	-32	23	-31	4.10	621

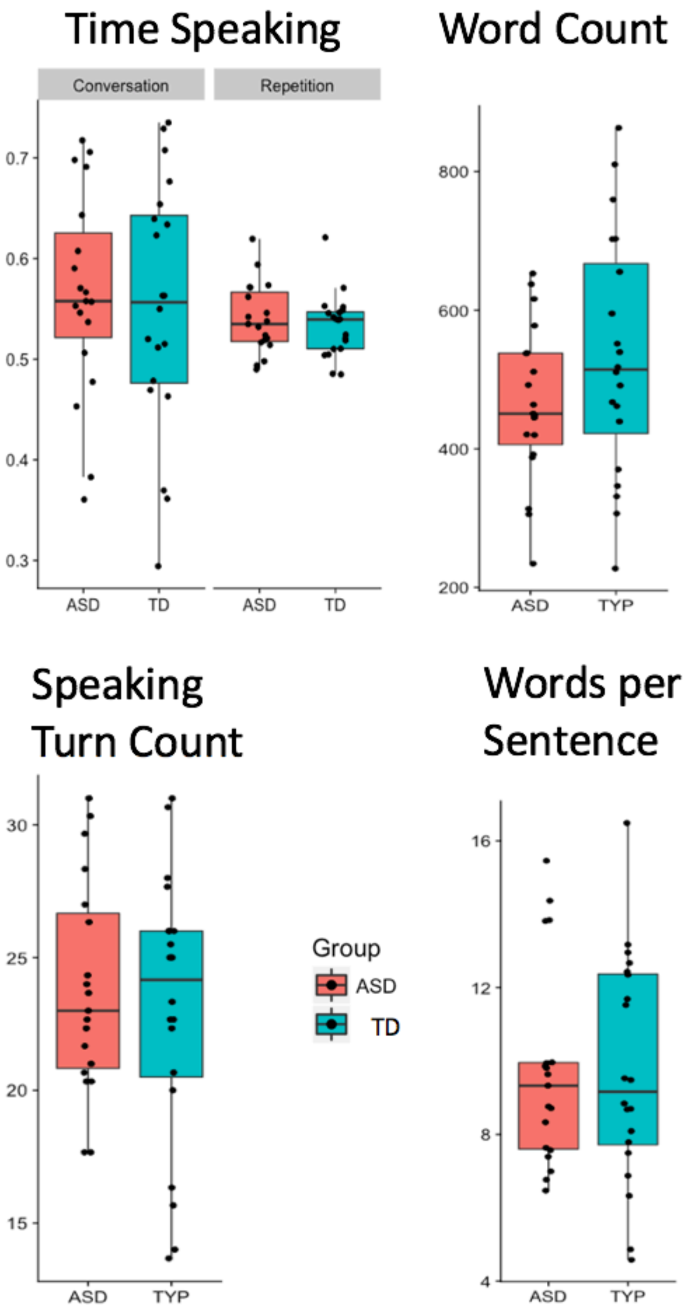
Supplementary Table 5: Regions determined using seed from Group main effect

	Center of Mass Coordinates			
Location	**x**	**y**	**z**	**mm3**
B/L medial somatosensory	-3	-40	45	2916
L fusiform/middle temporal gyrus	-43	-57	-15	2565
B/L lingual gyrus	-14	-50	-5	2106
B/L middle cingulate	1	-11	40	1971
B/L superior frontal gyrus	1	12	57	1917
L thalamus	-8	-12	11	1566
L precentral gyrus	-35	-32	50	1512
R lingual gyrus	10	-40	3	1485
R precuneus	9	-50	56	945
R anterior superior temporal gyrus/sulcus	46	-16	-7	918
R anterior superior temporal sulcus	51	2	-19	864
R ventral striatum	16	3	-9	783
R inferior frontal gyrus	44	4	6	702
R superior frontal gyrus	27	7	51	648
L temporal pole	-36	19	-29	621
L cerebellum	-14	-74	-40	567

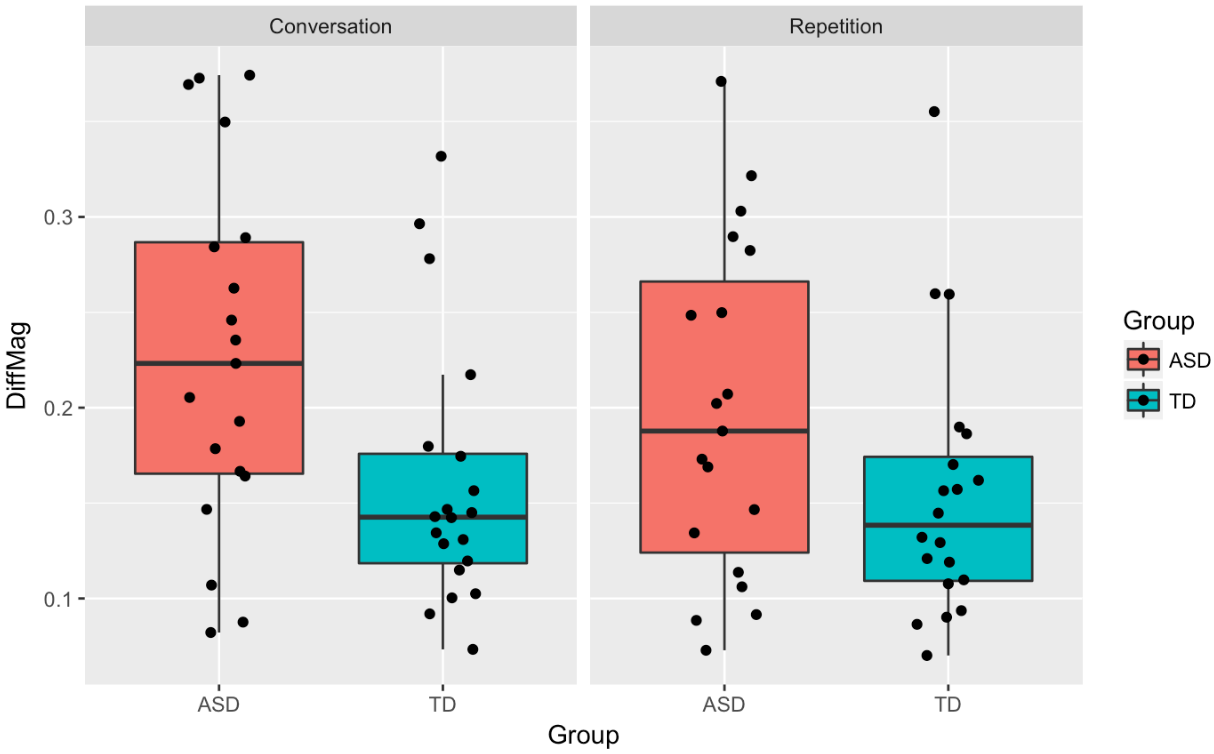
Supplementary Table 6: Regions determined using seed from Group X Task interaction (right fusiform gyrus)

Location	Center of Mass Coordinates				
	x	y	z	Peak *F*	mm^3
R precentral gyrus	41	-15	47	41.31	5103
R thalamus	6	-13	-5	29.97	3645

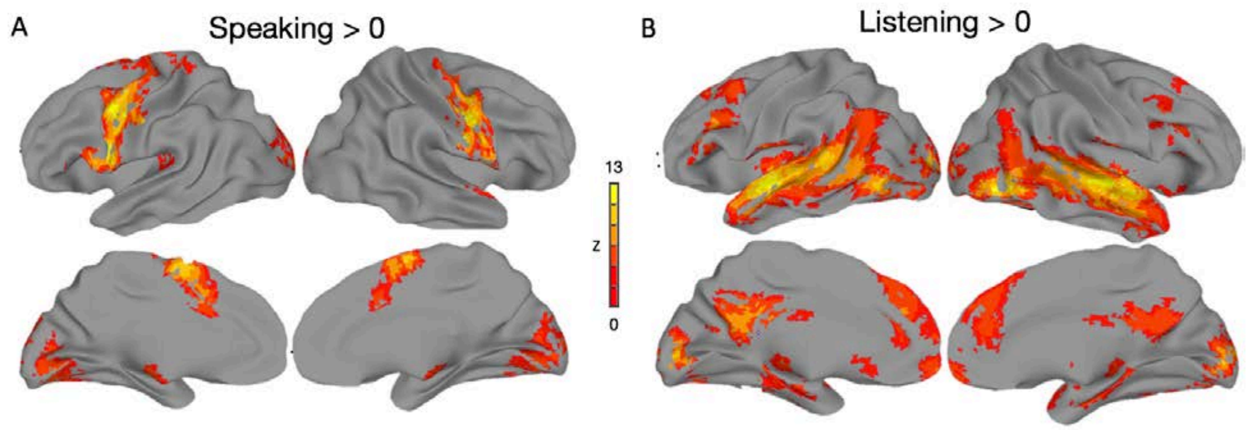
R posterior superior temporal sulcus	52	-43	7	23.52	2997
L calcarine gyrus	-13	-56	7	28.71	1404
L middle temporal area (MT)	-40	-63	10	27.25	1188
R extrastriate cortex	36	-79	28	26.1	1188
R inferior frontal gyrus (opercular)	50	14	13	21.4	1161
L posterior superior temporal sulcus	-52	-42	6	21.96	1134
L subgenual anterior cingulate	-4	15	-10	29.39	1026
L temporal pole	-52	8	-12	35.36	945
L declive	-30	-63	-20	23.6	918
Brain stem (pons)	-1	-29	-32	18.69	621
L pyramis	-7	-67	-23	22.55	567
R ventro-medial prefrontal cortex	3	62	-5	24.21	567



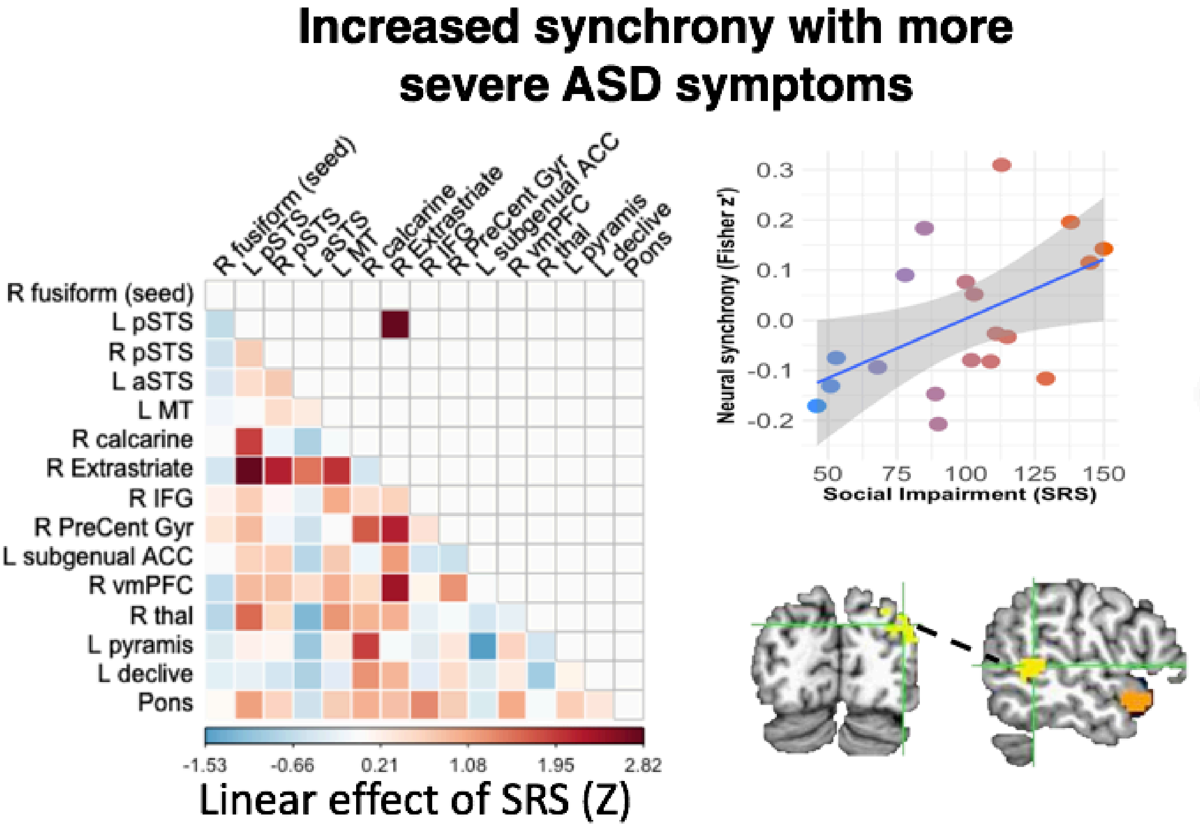
Supplementary Figure 1: Measures of verbal output. Autism (ASD) and control (TD) participants' verbal behavior was similar according to 4 basic measures. Both spoke for similar lengths of time relative to the experimenter (TimeSpeaking), produced similar numbers of word, similar numbers of turns, and sentences of similar lengths. Box and whisker plots show the median, with hinges indicating first and third quartiles, and points exceeding whiskers are 1.5* Inter-Quartile-Range.



Supplementary Figure 2: Motion Index averaged over runs for each subject, plotted by Group and Condition. The Motion Index (AFNI's @1dDiffMag) values used as nuisance covariates in the Conversation and Repetition conditions are shown for individual participants (black dots). Box and whisker plots show the median, with hinges indicating first and third quartiles and points exceeding whiskers are 1.5* Inter-Quartile-Range.



Supplementary Figure 3: Activity during Speaking and Listening turns. Audio recordings of the conversation were used to create regressor time series associated with speech production (Speaking) and speech comprehension (Listening) turns. These regressors were convolved with a hemodynamic response function and tested against a null hypothesis of 0 in a general linear model (which included task-driven activity). A) Producing speech was associated with increased activity in motor, premotor, left posterior superior temporal, left inferior frontal gyrus activity and supplementary motor cortex. B) Hearing speech was linked with activity in temporal, occipital, medial prefrontal and precuneal cortex. Height threshold of $p<.001$, cluster corrected to $p<.05$ over all voxelwise tests.



Supplementary Figure 4: Region-by-region functional connectivity for autistic participants during conversation modulated by Social Responsiveness Scale (SRS) in the Group-by-Task Interaction regions. Left posterior STS and right extrastriate cortex showed increased functional connectivity in participants with more severe autistic symptoms.

Adjustment and normalization of correlation coefficients

Where appropriate, to account for the different number of data points and degrees of freedom removed during the regressions, the correlation coefficients indexing functional connectivity were converted to population level estimates using the following formula (Supplemental Figure 5; Olkin & Pratt, 1958) before transformation to Fisher's z'. See also Gotts et al. (2012; Supplementary Materials).

$$\hat{\rho} = r\left(1 + \frac{1 - r^2}{2(df - 3)}\right)$$

Supplemental Figure 5: Formula adjusting sample correlation coefficient to the population estimate rho. (Olkin & Pratt, 1958) *r* = sample correlation coefficient. *df* = remaining degrees of freedom (# of fMRI time points - # of degrees of freedom removed during de-noising regression).