Large-Scale Assessment of Language, Speech, and Movement in Autism and ADHD with AI

Aimar Silvan[1], Lucas C. Parra[1]*, Adriana Di Martino[2], Michael P. Milham[2], Jens Madsen[1]

* corresponding author: parra@ccny.cuny.edu

[1]Department of Biomedical Engineering, City College of New York, New York, NY, 10031, USA
[2]Child Mind Institute, New York, NY, 10022, USA

Abstract

Attention-Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) frequently co-occur with overlapping symptoms. We investigated whether automated behavioral analysis during a clinician-child interview can identify distinct, objective features of the two disorders. Analyzing audio-video recordings of 2,341 youths (ages 5–22) in a broad community sample, multivariate models revealed that language difficulties often attributed to ADHD are primarily explained by age, cognitive ability, or co-occurring ASD. Increased motor activity specifically marked hyperactive-impulsive ADHD, but not ASD or inattentive ADHD. ASD was uniquely characterized by divergent narrative production and perspective-taking, alongside a distinct vocal profile of higher pitch and intensity, despite structurally intact language. While these digital behavioral measures correlate with most diagnostic categories and age, the joint analysis effectively separates the effects of ASD from ADHD. These findings show that scalable digital assessment from recorded clinical interviews can disentangle overlapping ASD and ADHD diagnoses into domain-specific behavioral signatures.

Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) are prevalent conditions with long-term functional impacts [1–4]. ADHD involves persistent inattention and/or hyperactivity, while ASD is characterized by challenges with social interaction and restricted, repetitive behaviors and interests [5]. These heterogeneous conditions frequently co-occur [6–12], sharing genetic factors [13–16] and converging variations in large-scale brain networks [17–22].

Current diagnostic assessments for ASD and ADHD rely on clinical observation and reports. The clinical visit is an opportunity to observe and assess patients' behaviors in a social context. Here, we test whether large-scale, automated analysis of naturalistic clinical interactions can disentangle overlapping ADHD and ASD diagnosis into separable behavioral dimensions.

We leveraged the Healthy Brain Network (HBN) cohort [23], a large community sample of neurotypical and neurodivergent youth. The data included a video-recorded, semi-structured interview where participants were asked about an emotive animated film they just watched. This easily scalable interview was designed to elicit language, emotional expression, and social-cognitive functions like perspective-taking.

State-of-the-art AI enables automated assessment of these audio-video recordings across three domains: gross body movements, language use, and manner of speech. In the following, we review relevant literature in these domains in the context of ADHD and ASD.

Effective communication depends on both structural language (vocabulary, syntax) and pragmatic language skills (social application, topic coherence) [24]. Social communication deficits are a core diagnostic feature of ASD but not ADHD [5]. However, studies using informant-based assessments suggest a significant degree of overlap, with both groups demonstrating difficulties in both pragmatic and structural language [25–27].

Beyond fundamental language skills, perspective-taking is also essential for discourse and narrative comprehension and construction [28]. This is often described as Theory of Mind (ToM), i.e., the ability to infer the mental and emotional states of others [29]. Difficulties in perspective-taking have long been associated with ASD [30]. Studies using automated language analysis have identified repetitive speech and atypical narratives in ASD [28,31], and narrative disorganization in ADHD [32]. However, these methods typically evaluate fewer than 50 participants [34] and analyze conditions separately, limiting generalizability and obscuring shared versus distinct traits. We address these gaps by jointly modeling structural, pragmatic, and higher-order socio-cognitive abilities across both diagnoses to develop a detailed model of traits in both ADHD and ASD on an unprecedented scale.

Alongside communication deficits, atypical prosody is a clinical hallmark of ASD [31]. Recent systematic reviews link ASD to higher mean pitch and greater pitch variability [33], though findings regarding intensity, speech rate, and voice quality remain inconsistent due to methodological heterogeneity and small sample sizes [33,34]. Furthermore, the influence of co-occurring ADHD, which may also present distinct prosodic features [35,36], remains poorly understood. While automated vocal analysis shows diagnostic promise [37–39], constrained samples currently prevent the isolation of robust, condition-specific vocal signatures for either group.

Motor differences are central to both ADHD and ASD, yet their behavioral signatures are rarely quantified under comparable conditions. While ASD has been associated with diverse motor patterns [40,41], it remains unclear which of these reflect distinct motor phenotypes or are shared manifestations of ADHD.

Further complicating matters, language, speech, and motor coordination develop rapidly with age, and symptom presentations can shift [43]. For instance, hyperactivity in ADHD often decreases while inattention persists into adolescence [44]. Without rigorously modeling key covariates like age, sex [45,46], and cognitive ability (IQ) [47,48], observed markers may simply reflect typical developmental or demographic variance rather than true psychopathology.

This first analysis of the HBN semi-structured interview presents an opportunity to simultaneously observe these diverse behavioral domains at an unprecedented scale. Using

gold-standard clinical diagnoses, we apply a multivariate approach to identify independent drivers of behavioral variance. Deviating from prior work that treats co-occurring ADHD and ASD as a distinct categorical group [49–51], we model ADHD Inattention, ADHD Hyperactivity, and ASD as independent clinical dimensions. By mapping digital phenotypes onto partially overlapping diagnostic dimensions rather than mutually exclusive categories, our findings help reconcile prior inconsistencies in the literature, and isolate condition-specific behavioral signatures.

Results

A semi-structured interview of a community sample for automatic behavioral analysis

To measure naturalistic behaviors, 2,341 participants (aged 5-22) of the HBN cohort were video-recorded during a semi-structured interview after watching "The Present", a brief animated film in which a boy's interaction with a three-legged puppy culminates in the revelation that both share a missing leg. During the interview, participants were asked identical questions one at a time (Table S1) probing narrative recall, factual memory, emotional interpretation, and perspective-taking. We used state-of-the-art computational tools to derive objective, quantitative metrics from the video-recorded interviews, covering expressive language, vocal prosody, as well as facial and body movements (Fig. 1A).

The cohort was diagnostically rich and clinically complex (Fig. 1B, C & Table 1). The sample exhibited an overall male-to-female ratio of approximately 2:1 (Fig. 1D), with lower than chance prevalence in females (17% of the ASD group and 23-27% of the ADHD group; Fig. 1E). Consistent with the clinical literature [52], a clinician-confirmed diagnostic status of either ADHD-Inattentive or ADHD-Hyperactive significantly increased the odds of co-occurring ASD (Adjusted Odds Ratios (OR) = 2.3 and 1.6, respectively, Fig. 1F). Clinical overlap extended to instrument ratings, which were highly correlated and strongly influenced by age and sex (Fig. 1G). For instance, parent-rated hyperactivity traits decreased significantly with age ($\rho(2194) = -0.28$, $p<0.001$, Bonferroni corrected, Fig. 1H), consistent with typical developmental trajectories [44]. The same was also reflected in the effect of age on the prevalence of ADHD-Hyperactive status (Fig. 1F, OR = 0.83). Such a change with development is to be expected for language abilities [53], which are a primary focus of this study.

Note that in Fig. 1E-F and in all our analyses, ADHD-Combined presentation is coded as both Inattentive and Hyperactive. To distinguish this coding with two binary variables (Inattentive, Hyperactive) from the three DSM-5 ADHD presentations (Inattentive, Hyperactive, Combined), we will refer to these as diagnostic *status* and *presentation*, respectively.

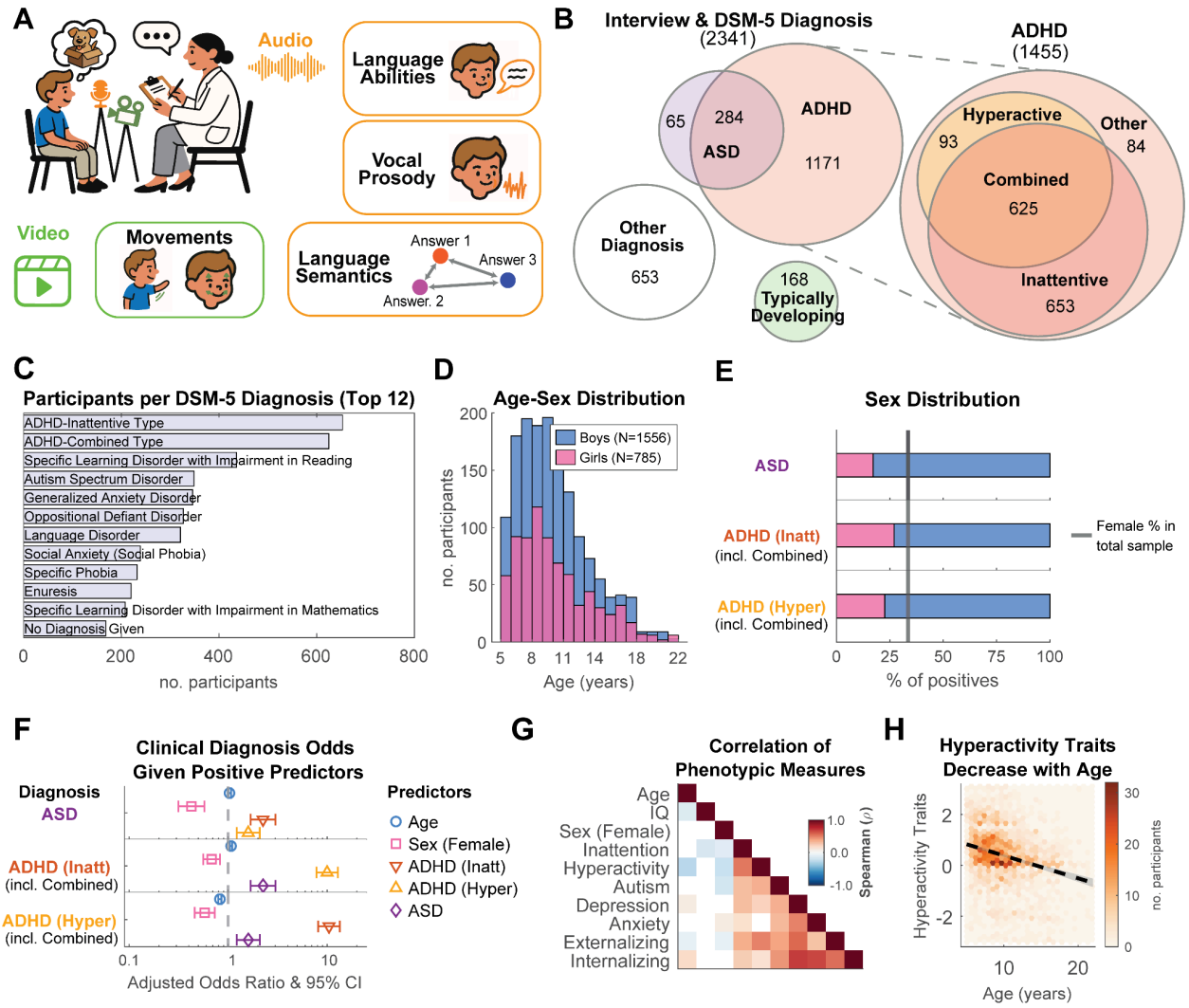

Figure 1. Overview of the study cohort. **A)** Using AI for objective behavioral characterization. Video recordings were processed to quantify body and facial movements, while audio was transcribed and analyzed to extract metrics of language structure, semantic content, and vocal prosody. **B)** Venn diagram showing the distribution and overlap of the primary diagnostic groups. **C)** The 12 most common clinician-assigned DSM-5 diagnoses in the full cohort (N=2,341), see Fig. S1 for specific co-occurring diagnoses for the ADHD and ASD children. **D)** Age and sex distribution of the full cohort, see Fig. S1 for age-sex distributions per diagnostic group in panel B. **E)** Sex distribution within the children with ASD and/or ADHD diagnostic, ADHD-Combined are here considered positive in both hyperactive and inattentive diagnostic status. The black line indicates chance occurrence based on the sex ratio in the cohort. **F)** Adjusted odds ratios showing the likelihood of a co-occurring positive diagnosis in individuals with positive specific predictors (or one-year change in the case of age). Females are less likely than males to receive an ADHD or ASD diagnosis (OR < 1). **G)** Spearman correlation matrix, displaying only significant associations ($p<0.01$, Bonferroni corrected, white otherwise). The matrix illustrates that clinical instruments for specific neurodevelopmental traits, like ASSQ for autism and SWAN for ADHD, are highly correlated with other measurements like depression (MFQ), anxiety (SCARED), and internalizing/externalizing behaviors (CBCL). They also correlate with age, IQ, and sex, which complicates the attribution of observed behavior to any single instrument scale independently. **H)** Two-dimensional histogram showing the negative correlation between age and hyperactivity traits (SWAN

scale) (ρ(2194) = -0.28, p<0.001). Full names for clinical instruments (SWAN, ASSQ, MFQ, SCARED, CBCL) are provided in the 'Clinical Instruments' section in the Methods. Detailed model statistics for panel F are available in Supplementary Section S1.1.

Table 1. Demographic, clinical, and interview characteristics

Characteristic		ADHD-Inatt[b] (N=1278)	ADHD-Hyper[b] (N=718)	ASD(ADHD+)[c] (N=284)	ASD(ADHD-)[c] (N=65)	Other Diagnoses (N=653)	Typically Developing (N=168)
Age (in years), mean (SD)		10.1 (3.2)*	9.1 (2.9)	10.2 (3.5)	10.6 (4.6)	10.6 (3.6)**	9.4 (3.2)
Sex, N (%)	Female	346 (27)**	162 (23)**	40 (14)**†	20 (31)†	291 (45)	74 (44)
	Male	932 (73)	556 (77)	244 (86)	45 (69)	362 (55)	94 (56)
Co-occurring Diagnosis, N (%)	ADHD-Inatt[b]	N/A	625 (87)	264 (93)	N/A	N/A	N/A
	ADHD-Hyper[b]	625 (49)	N/A	166 (58)	N/A	N/A	N/A
	ASD	264 (21)	166 (23)	N/A	N/A	N/A	N/A
Clinical Instruments, mean (SD)	Inattention (SWAN)	1.2 (0.9)**	1.1 (0.9)**	1.2 (0.9)**††	0.5 (1.1)**††	0.0 (1.0)	-0.3 (1.0)
	Hyperactivity (SWAN)	0.7 (1.0)**	1.2 (0.8)**	0.9 (0.9)**††	0.3 (1.0)**††	-0.3 (1.0)	-0.3 (1.0)
	Autism (ASSQ)	8.5 (9.2)**	10.2 (10.1)**	17.4 (10.9)**	18.1 (11.0)**	4.1 (5.9)	2.3 (3.1)
	Depression (MFQ)	10.0 (8.8)**	9.8 (8.0)**	10.5 (8.8)**	9.4 (8.6)**	7.8 (8.4)**	4.7 (4.9)
	Anxiety (SCARED)	14.7 (12.1)**	14.6 (12.0)**	15.8 (12.2)**	19.9 (15.3)**	14.4 (11.7)**	9.4 (8.0)
	Externalizing (CBCL)	12.5 (9.8)**	15.4 (9.8)**	12.8 (9.7)**	10.2 (7.7)**	8.3 (8.3)*	5.6 (5.7)
	Internalizing (CBCL)	10.5 (8.3)**	10.4 (8.0)**	12.1 (8.1)**	12.9 (8.5)**	9.6 (8.4)**	5.3 (5.0)
IQ [a], mean (SD)	Full-Scale	98.5 (16.3)**	100.3 (16.2)**	96.7 (18.5)**	96.9 (20.7)**	98.5 (16.9)**	106.5 (14.3)
	Verbal	103.2 (16.3)**	104.4 (16.1)*	100.7 (19.1)**	99.7 (22.1)**	101.7 (16.7)**	109.4 (13.8)
Interview length (minutes), mean (SD)		4.5 (1.5)	4.5 (1.5)	4.6 (1.7)	4.7 (1.4)	4.5 (1.4)	4.5 (1.9)

[a] IQ was measured using WISC-V [54], with exceptions for early study participants using WASI [55], and children under 6 or with an IQ below 70 using KBIT [56]. See the 'IQ Score' section in Methods for details. [b] ADHD-Combined are here considered positive in both hyperactive and inattentive diagnostic status. [c] 'ASD(ADHD+)' are ASD participants with any co-occurring DSM-5 ADHD diagnosis, 'ASD(ADHD-)' without. [d] 'Other Diagnoses' includes participants with one or more positive DSM-5 diagnoses, but negative ADHD and ASD diagnoses.

Statistical differences for continuous measures were assessed using independent samples t-tests (for normally distributed data) or Wilcoxon rank-sum tests (for non-normally distributed data), following normality checks via the Kolmogorov-Smirnov test. Categorical comparisons were performed using Chi-square tests. * $p < 0.01$, ** $p < 0.001$ compared to the Typically Developing group (Bonferroni corrected for 5 comparisons). † $p < 0.01$, †† $p < 0.001$ for the comparison between ASD(ADHD+) and ASD(ADHD-). Detailed statistical reports for each comparison are available at Table S2. Full names for clinical instruments (SWAN, ASSQ, MFQ, SCARED, CBCL) are provided in the 'Clinical Instruments' section in the Methods.

Developmental factors dominate over diagnostic effects in structural language

To characterize naturalistic communication at scale, we processed the audio recordings of the interview using a state-of-the-art AI pipeline that integrates automatic transcription with Large Language Model (LLM)-based diarization. From these transcripts, we extracted objective metrics of structural and pragmatic language abilities (Fig. 2A).

To isolate the unique behavioral contributions of each diagnosis while accounting for high co-occurrence and developmental maturation, we employed a multivariate regression model. This approach coded diagnostic status as independent binary variables (ASD, ADHD-Inattentive, ADHD-Hyperactive) while simultaneously controlling for Age, Sex, and IQ (Fig. 2C).

This multivariate approach demonstrated that structural language abilities are overwhelmingly driven by development, not by ADHD or ASD. For example, a simple univariate analysis suggested a significant negative association between lexical diversity and hyperactivity ($\rho(2160)=-0.16$, $p<0.001$, Bonferroni Corrected, Fig. 2B). However, the multivariate model revealed this variation was actually driven by the effect of Age (Effect Size=0.489; effect sizes (ES) reported in this study represent Cohen's d for binary and Cohen's f for continuous variables; see Methods) and IQ (ES=0.273) (Fig. 2C-D). This pattern held across almost all structural language metrics (Fig. 2E). Older participants spoke more (ES=0.389) and faster (ES=0.563), while females spoke faster (ES=0.165) and answered more coherently (ES=0.189) than males. The only significant diagnostic marker identified in basic language use was an increase in self-referential speech (first-person pronouns) specific to participants with ASD (ES=0.278).

Our planned analysis was to model ADHD and ASD as independent additive effects. The alternative approach of modeling the co-occurrence as a separate diagnostic category (ADHD+ASD) [49–51] is treated in the Methods under "Alternative analysis approaches" along with Age interactions and other post hoc analyses.

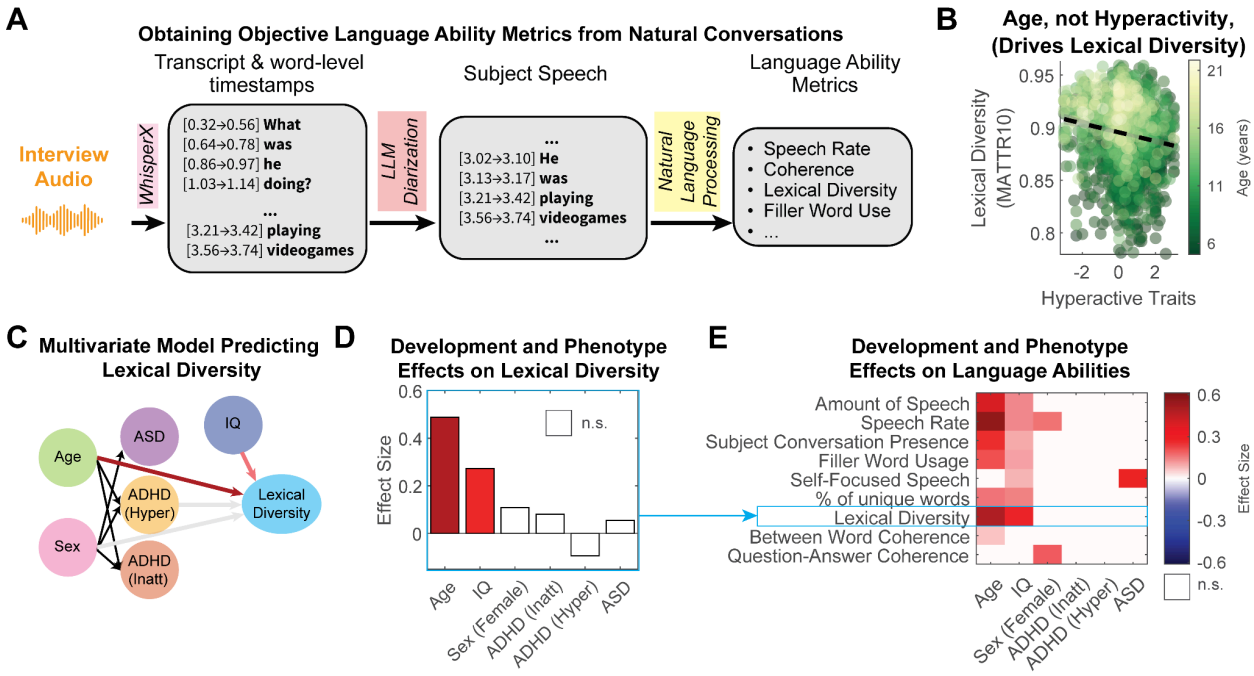

Figure 2. Language abilities are primarily driven by development, not by ADHD or ASD. A) The computational pipeline used to extract objective structural and pragmatic language metrics from transcribed audio. **B)** Hyperactive traits (SWAN) and lexical diversity (Moving-Average Type-Token-Ratio with a window of 10) are correlated ($\rho(2160)=-0.16$, $p<0.001$). However, the color gradient, representing age, reveals that this association is confounded by development. **C)** Conceptual diagram of the multivariate regression model used to isolate the unique effects of each predictor. A "Univariate Model" does not account for covariation, while the "Multivariate Model" controls for shared variance, with Age and Sex as a common source of variance. Arrows are color-coded to the effect size magnitude in the multivariate regression model (panel D). Univariate associations between DSM-5 diagnoses and demographics with lexical diversity (grayed arrows) are instead better explained by age and IQ in the multivariate model. Black arrows indicate effects tested in Fig. 1F. (See Supplementary Section S1.1). **D)** Effect sizes from the multivariate model predicting lexical diversity. Effect sizes represent the unique contribution of each factor (Sex coded as Female = 1, Male = 0). For binary variables, the reported effects are Cohen's d, while for numerical variables, we report Cohen's f (see Methods). **E)** A summary matrix of effect sizes from the multivariate model applied to a range of language metrics. The results show the consistent and strong effects of age and IQ, and the general lack of diagnostic effects across all measured aspects of language abilities. Significance cut-off at $p < 0.01$ (white otherwise; Bonferroni-corrected across the 9 models tested). Detailed multivariate model reports are available in Supplementary Section S1.2.

ASD is uniquely associated with atypical narrative, story comprehension, and perspective taking

While structural language remained largely intact across diagnoses, impairments in social communication often stem from higher-order difficulties in narrative construction and perspective-taking (Theory of Mind). To objectively quantify these functions, we computationally compared the semantic content of each participant's interview responses to those of their age-matched, Typically Developing peers, generating an objective measure of "answer typicality" (Fig. 3A-C).

After controlling for Age, our models revealed a significant, divergent socio-cognitive profile specific to ASD (Fig. 3E). An ASD diagnosis was uniquely associated with producing less typical narratives (ES=-0.299, Fig. 3D), showing different thematic understanding -- such as failing to convey that the boy and the puppy were similar because they were both missing a leg (Fig. 3F) -- (ES=-0.266), providing idiosyncratic descriptions of their own emotions (ES=-0.272), and expressing atypical preferences (ES=-0.313). In contrast, neither the ADHD-Inattentive nor ADHD-Hyperactive statuses were significantly associated with any of these social-cognitive domains.

Despite measuring similarity of responses relative to age-matched controls, age remained a powerful predictor across some semantic language use measures (Fig. 3E). This pattern is consistent with increasing convergence towards age-normative responses, rather than a drift in the semantics of the responses. Indeed, the exemplary ‘typical answers’ for each age bin were remarkably similar (Table S1), suggesting no age-related difficulties in understanding or answering the questions within the Typically Developing population. Consistent with social-cognitive maturation [43], older participants provided more consistent, detailed narratives (ES=0.394) and recalled the movie's main message more typically (ES=0.345). Similarly, higher IQ was associated with more typical narrative abilities (ES=0.185), thematic understanding (ES=0.212), and factual memory (ES=0.153). We also observed a secondary finding related to sex, with females tending to produce more typical narratives (ES=0.223), and positive and negative preferences (ES=0.264 & 0.180) compared to male participants (Fig. 3E).

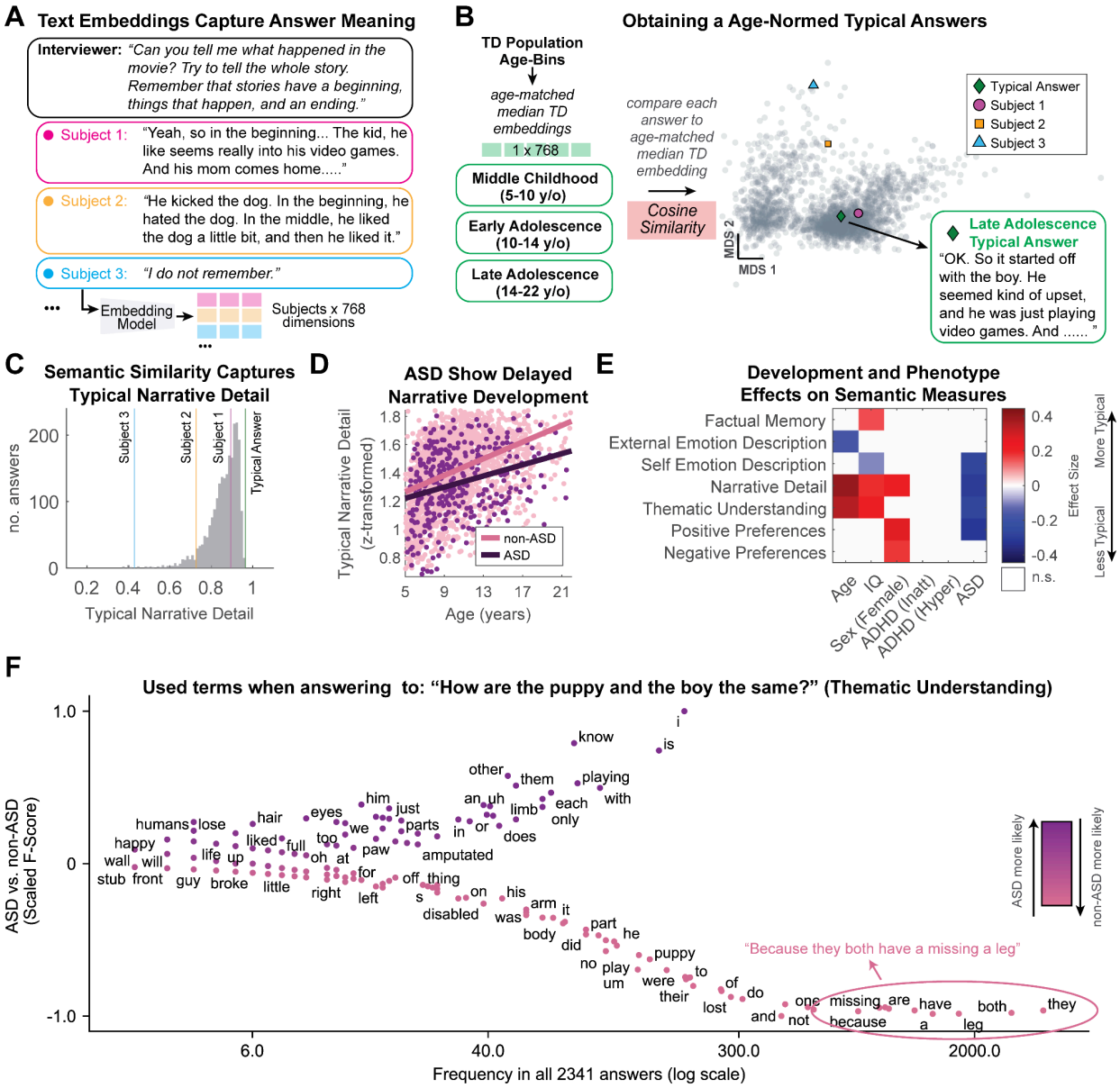

Figure 3. Semantic analysis reveals ASD-specific effects on narrative and social-cognitive abilities. A) The computational pipeline for semantic analysis. Participants' answers to 23 predefined interview questions were parsed and converted into numerical semantic representations (embeddings). **B)** Multidimensional Scaling (MDS) representation of the answer embeddings, where similar embeddings are represented closer together. A "typical answer" is obtained from the semantic (cosine) similarity to the median embedding of the corresponding age-matched TD group answers. **C)** Semantic similarity to the age-appropriate prototype is used as a measure of 'typicality'. For the narrative recall question in panel A, higher similarity reflects a more detailed and developmentally expected recount of the narrative relative to age-matched peers. Subject 1 provides a detailed answer (high similarity), whereas Subject 3 does not recall the story (low similarity). **D)** Joint distribution illustrating the positive association between age and the production of more typical narratives, but the delayed development in the ASD group compared to the non-ASD group. **E)** Matrix of effect sizes from the multivariate model, assessing the impact of diagnosis and demographic factors on various social-cognitive domains probed by the interview questions (e.g., factual memory, emotional description, and thematic understanding). Only effect sizes with p<0.01, after

Bonferroni correction, are shown in color. See Table S1 for a full list of questions and the typical answers for each age bin, and Supplementary Section S1.3 for detailed multivariate model reports. **F)** Scatterplot of terms distinguishing ASD (top, purple) from non-ASD (bottom, pink) participants in responses to the question related to thematic understanding, using *scattertext* [57] (See Methods section 'Visualization of Distinctive Terms' for details, and Fig. S2 for visualization of other questions).

Autism Spectrum Disorder has a distinct vocal profile

The acoustic patterns of pitch, intensity, and rhythm are critical for conveying the intent of an utterance. To this end, we employed an automated pipeline to extract a comprehensive set of acoustic features from the segmented speech of each child, including measures of pitch, loudness, and voice quality (Fig. 4A).

Our automated acoustic analysis revealed a distinct vocal signature specific to ASD. Controlling for covariates, ASD was significantly associated with increased mean pitch (ES=0.222), greater pitch variability (ES=0.261), and higher vocal loudness (ES=0.231) (Fig. 4C). Furthermore, participants with ASD exhibited higher dysphonia (ES=0.371, Fig. 4B) and increased breathiness (ES=0.325). Notably, this vocal profile was highly specific: neither ADHD status was associated with altered vocal prosody. Age again proved to be a dominant predictor, with older participants naturally exhibiting reduced pitch (ES=-0.419), lower pitch/loudness variability (ES=-0.415 & -0.140), and decreased breathiness (ES=-0.104).

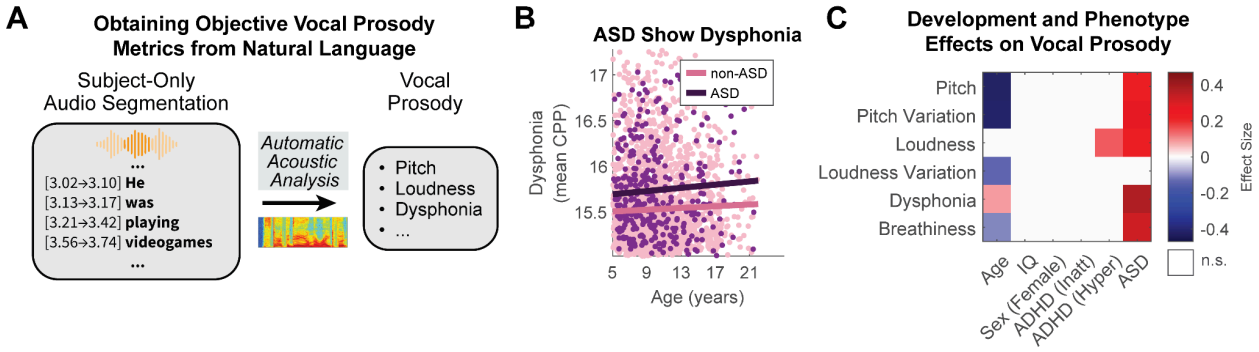

Figure 4. Automatic vocal prosody metrics reveal an ASD-specific acoustic profile. A) Automatic extraction of vocal prosody features involved transcribing and diarizing speech, thus isolating the children's speech from that of the adult interviewer, and extracting audio-based features using automatic acoustic analysis.[58] **B)** ASD participants showed increased dysphonia compared to non-ASD participants, which may be related to an "unusual" voice in ASD. **C)** Effect Size matrix illustrating the effects of development and ASD on vocal prosody. Effect Sizes in this matrix are estimated again using a multivariate regression model, with $p<0.01$, after Bonferroni correction. See Supplementary Section S1.4 for detailed multivariate model reports.

Increased motor activity is uniquely associated with ADHD-Hyperactive diagnostic status

Motor behaviors are relevant diagnostic features of ASD and ADHD, with hyperactivity being a key diagnostic criterion in ADHD, and stereotypical and repetitive motor behaviors being among the defining criteria for ASD. Yet motor behaviors are rarely quantified objectively in naturalistic settings. Here, we applied computer vision models to extract objective frame-to-frame

displacement metrics for the upper body, head, eyes, and face (Fig. 5A). Consistent with typical maturation, age was a strong negative predictor of movement categories (ES ranging from -0.461 to -0.308).

Once developmental decreases in movement were accounted for, we identified a highly specific motor signature: ADHD-Hyperactive status was robustly associated with increased movement across all measured body and facial regions (ES=0.225-0.301) (Fig. 5B-C). Conversely, neither ASD nor ADHD-Inattentive status were linked to significant changes in motor activity, confirming that objective video-derived kinematics isolate the hyperactive-impulsive domain of ADHD without generalizing to the inattentive presentation or ASD.

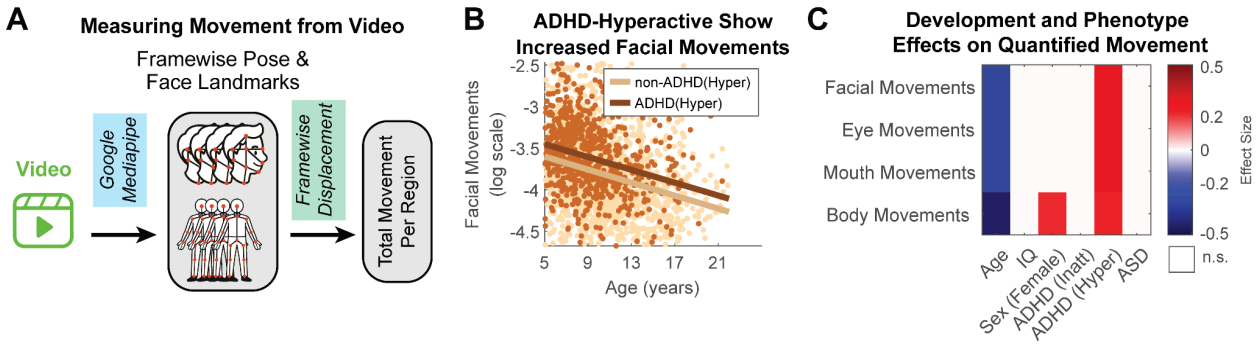

Figure 5. Associations of demographics and phenotype with body and face movements. A) We use Google's Mediapipe computer vision models to obtain framewise locations for anatomical and facial landmarks. We then calculate the average framewise displacement for the different regions of interest. **B)** ADHD-Hyperactive youth show increased facial movements across the whole developmental spectrum, compared to non-ADHD Hyperactive. **C)** Effect Size matrix illustrating the effects of development and an ADHD-Hyperactive status on all measures of face and body movements. Effect Sizes in this matrix are estimated using multivariate regression, with $p<0.01$, after Bonferroni correction. See Supplementary Section S1.5 for detailed multivariate model reports.

Discussion

This study applied automated digital phenotyping to a large community sample to isolate behavioral dimensions of ASD and ADHD, demonstrating that while narrative divergence and atypical speech uniquely characterize ASD, increased motor activity specifically marks ADHD-Hyperactivity. Crucially, language deficits often attributed to ADHD are fully explained by age, IQ, and co-occurring ASD.

Previous standardized assessments suggest ADHD involves structural and pragmatic language impairments [59,60], with community prevalence estimated at 40-45% [61,62]. However, standardized tests impose heavy demands on executive functions like sustained attention and working memory. Consistent with other naturalistic studies [63] our conversational analysis shows that once age, IQ, and ASD are controlled, structural, conversational language in ADHD is intact. This extends to higher-order social cognition. After accounting for age, IQ, and ASD status, ADHD was not independently associated with narrative ability or perspective-taking (ToM), indicating that previously reported differences may reflect developmental factors or diagnostic overlap rather than ADHD-specific effects reported previously [64,65]. Apparent communication

difficulties in ADHD likely reflect executive dysfunction, such as poor inhibition, impacting speech organization, rather than fundamental deficits in linguistic comprehension or mentalizing [66,67]. Impaired ToM thus remains a critical differentiator indicating co-occurring ASD in youths with ADHD [68].

Conversely, our ASD cohort exhibited mostly typical structural language trajectories, supporting classic conceptualizations of the disorder as a core social communication deficit [5]. The primary structural deviation was increased self-referential speech via first-person pronouns, reflecting an asynchrony between linguistic and social development [69]. ASD communication deficits emerged strictly in higher-order domains: participants produced less typical narratives, missed overarching thematic 'gists' [70], and provided idiosyncratic emotion descriptions [71]. This naturalistic validation aligns with computational language analyses demonstrating diminished narrative coherence in demanding or emotive contexts for ASD [72,73].

Furthermore, we objectively quantified the long-observed atypical prosody of ASD [34]. Our analysis confirmed higher pitch and pitch variability [33], while clarifying past inconsistencies by identifying increased vocal intensity, dysphonia, and breathiness as distinct, ADHD-independent ASD markers, validating clinical descriptions of "hoarse" voices [74]. Most importantly, these associations were absent in ADHD.

Regarding motor behavior, unobtrusive computer vision effectively captured ADHD-related restlessness that typically requires specialized sensors [42,75]. Accounting for age-related decreases, widespread head, body, and facial movement specifically marked the ADHD-Hyperactive status. Neither ADHD-Inattentive nor ASD diagnoses affected the global displacement metrics we used here. Because motor challenges are core to ASD,[41] the absence of an ASD-associated signal in these global kinematic measures suggest that they may not capture the more specific, repetitive movement patterns often described in autism [40] which may require a more granular movement classification [40,76–81].

Developmental and demographic factors proved essential to the behavioral phenotype, with age robustly driving changes across motor, linguistic, and socio-cognitive domains alongside sex-specific profiles. Digital biomarkers must rigorously disentangle normative trajectories from pathology. Although cognitive ability varies within ASD, [47,48] and influences language [31,82], and IQ remains controversial as a pure covariate [83], our primary diagnostic associations remained robust whether Full-Scale or Verbal IQ were included or excluded (see Fig. S3).

While our automated features correlate strongly with established clinical instruments (see Fig. S4), they can offer enhanced presentation-specificity. However, these computational tools are not intended to replace current diagnostic methods, as traditional instruments still align more closely with categorical DSM-5 definitions (see Fig. S5).

As a cross-sectional study, we could identify developmental trends but could not establish causality or track individual trajectories over time. Furthermore, our cohort primarily included individuals with mild-to-moderate ASD traits (see Fig. S6). This likely explains why manual

inspections revealed no overt behavioral stereotypies, our analysis focused on global kinematics instead.

Future work on ADHD and ASD should validate these brief, video-based measures longitudinally using wearable sensors and apply them within standardized assessments like the ADOS [84]. More broadly, this work illustrates how brief, standardized clinical interactions can be transformed into multidimensional behavioral measures at scale. A similar approach may be relevant beyond developmental disorders, to include neurological and psychiatric conditions where patients present with altered movement, posture, speech or language.

Methods

Clinical Population

This study used video-recorded conversations with 2,341 children (aged 5-22, M=10.1, SD = 3.39 years) as part of the Healthy Brain Network (HBN) [23]. The cohort included 1,455 participants with clinician-confirmed ADHD: 653 with inattentive presentation, 93 with hyperactive-impulsive presentation, and 625 with combined presentation. An additional 84 showed ADHD traits without meeting diagnostic criteria for a specific ADHD presentation (DSM-5 ADHD-other/unspecified). The sample also comprised 349 individuals with mild to moderate ASD (mostly level 1 according to DSM-5). Of the 1455 participants with ADHD, 284 (19.5%) had a co-occurring ASD diagnosis, a rate similar to those reported in previous community-based cohorts [85,86]. The cohort included 168 Typically Developing individuals and 653 with other clinical diagnoses, mostly anxiety and learning disorders. This group did not exhibit significant differences in autistic, hyperactive, or inattentive traits compared to the Typically Developing controls (Table 1), supporting their distinction from the primary target groups in specific ADHD and ASD traits. Furthermore, despite the wide developmental range of the study, there were no marked differences in age distributions between diagnostic and Typically Developing groups (Table 1 and Fig. S1). The overall male-to-female ratio was 1.98:1 (1556/785). Detailed demographic and clinical characteristics per diagnosis are available in Table 1.

Of the 2,341 participants, 1143 (48.8%) were White/Caucasian, 311 (13.3%) Black/African American, 226 Hispanic (9.7%), 55 (2.3%) Asian, and 42 (1.8%) identified as other races. 382 (16.3%) identified as two or more races. 182 (7.8%) participants did not specify, or their race was unknown.

Behavioral Task and Data Acquisition

Naturalistic behavior was recorded during a semi-structured clinical interview following the viewing of an emotive animated short film ("The Present"). The film depicts a boy who initially rejects a three-legged puppy given to him by his mom while playing a videogame; in the final scene, it is revealed that the boy himself has a missing leg. The narrative therefore invites viewers to infer the characters' emotional states and the similarity between them. The subsequent interview was structured to elicit responses related to narrative recall, emotional description, and perspective-taking (Theory of Mind) (see Table S1 for a list of the questions

and Table S3 for an overview of related measures). High-definition video (Canon XC15) of the participant's face and upper body and high-fidelity audio (Røde NT1) were captured simultaneously for automated analysis. Examples for the interview settings and camera view are available at Fig. S7.

Diagnostic Procedure

The HBN is a large-scale biobank that utilizes a community-referred recruitment model to capture the wide heterogeneity inherent in developmental psychopathology. As such, the HBN sample is enriched for individuals meeting DSM-5 criteria.

Diagnoses within the HBN are determined by licensed clinicians at the Child Mind Institute based on a consensus model that integrates multiple sources of information. The primary diagnostic instrument is the semi-structured Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) [87] for DSM-5, which is administered to both participants and their parents. The final clinician consensus diagnosis also incorporates behavioral observations during extensive testing, parental and self-reported clinical questionnaires, and the participant's developmental and educational history. For ASD, the primary evaluation relies on the KSADS autism module and questionnaires (e.g., SRS-2, SCQ, ASSQ). Participants with suspected autism received supplemental "gold-standard" assessments like the ADOS or ADI-R to confirm diagnosis by trained clinicians. Further details on the diagnostic process and clinician-administered assessments are available on the original HBN paper [23].

For our analysis, clinician-confirmed diagnoses were modeled to assess the impact of specific conditions. ADHD diagnoses were categorized as inattentive and/or hyperactive/impulsive diagnostic status. Participants with a clinician-confirmed diagnosis of ADHD-Combined presentation were considered positive for both hyperactive and inattentive. ASD diagnoses were modeled as a single binary variable (1 for positive, 0 for negative). The vast majority of these individuals (>95%) had mild to moderate presentations (only a few requiring significant support, DSM-5 level 1), consistent with the HBN's inclusion criteria, which require participants to be verbal (see Fig. S6 for distribution of autism instrument scores). The control group, designated as the Typically Developing Population, consisted of individuals with a confirmed negative diagnostic clinician decision in all DSM-5 categories.

Clinical Instruments

Participants in the HBN are deeply phenotyped using a comprehensive battery of clinical instruments. For this study, we defined ADHD traits using the Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale (SWAN)[88], and ASD traits using the Autism Spectrum Screening Questionnaire (ASSQ)[89]. These scales were selected as they showed the strongest unique association with their respective clinician-confirmed diagnoses compared to other available scales (see Fig. S5). Other clinical instruments used in this study include the Mood and Feelings Questionnaire (MFQ, depressive traits)[90], Screen for Child Anxiety Related Disorders (SCARED-P, anxiety traits)[91], and the Child Behavior Checklist (CBCL)[92], with two factors: internalizing and externalizing traits. A full description and statistical analysis of age, sex, IQ, and clinical instrument scores for each diagnostic group is provided in Tables 1 and S2.

IQ score

Cognitive ability was assessed using a composite IQ score. Full-Scale IQ scores were primarily measured using the Wechsler Intelligence Scale for Children (WISC-V) [54] (N=1984, mean +- SD = 99.2 +- 16.7). Exceptions included early participants who were administered the Wechsler Abbreviated Scale of Intelligence (WASI) [55] (N=36, 97.8 +- 14.7), and children under age 6 or with known IQ below 70, for whom the Kaufman Brief Intelligence Test (KBIT) [56] was used (N=166, 100.4 +- 15.8). Full-Scale IQ Composite scores were available for 2186 out of the 2341 participants (93.3%). Similarly, verbal IQ scores were obtained by combining the WISC-VCI (N=1986, 103.6 +- 16.6), WASI-VCI (N=36, 99.3 +- 14.4), and KBIT-verbal (N=166, 100.6 +- 15.9) scores. Verbal IQ scores were available for 2188/2341 participants (93.5%). The main analysis and results were computed using Full-Scale IQ as a covariate. However, the main associations between diagnoses and measured behavior were independent of including IQ as a covariate or not (Fig. S3). Similarly, the results did not change significantly if modeling with Verbal IQ scores instead.

Transcription, Diarization, & Question-Answer extraction

Audio recordings were transcribed using the WhisperX model (v.3.3.0, Whisper model *large-v2*)[93], generating a raw transcript with word-level timestamps. Validation against a manually transcribed and diarized ground-truth dataset (N=50, author AS) demonstrated a Word Error Rate (WER) of 3.75% +- 4.59% (M +- SD). To separate interviewer and participant speech, we utilized a custom prompt with Google's large language model (*gemini-2.0-flash*). This context-aware approach yielded a Diarization Error Rate (DER) of 3.16% +- 3.97%, significantly outperforming standard acoustic-based methods (*pyannote/speaker-diarization-3.1* [94] DER: 16.35% +- 8.78%). The LLM leveraged the semi-structured interview context to infer speaker turns, avoiding acoustic ambiguities caused by pitch and/or speech overlap between children and interviewers. A secondary LLM prompt subsequently isolated specific interview questions and the participant's corresponding answers. Complete prompts are provided in the Supplementary Sections S2 and S3. The number of valid answers per question is available in Fig. S8.

Description of Behavioral Variables

Various open-source tools were used to obtain objective measures of language, speech prosody, semantic content, and movement from the interview data. An in-depth description of the measures and their extraction methods is provided in Table S3; the distribution of the measures is available in Figure S9. The code for the extraction of measures is available under the link in the Code Availability section.

Language and Speech Prosody Measures. We extracted a set of acoustic and linguistic features using the Openwillis toolkit (v.3.0.5) [58]. Openwillis is a Python wrapper developed by Brooklyn Health that automates feature extraction by integrating previously established open-source speech and natural language processing tools for the extraction of objective language and speech measures. See Table S3 for a complete description of all measures and their extraction methods.

Age-Normed Semantic Measures. To characterize the semantic content of participants' answers, we utilized Google's 'Gecko' text embedding model (*text-embedding-004*) to generate 768-dimensional numerical representations of each response that capture meaning in a high-dimensional semantic space [95]. To distinguish diagnosis-related divergence from typical developmental immaturity, we adopted an age-normed approach. First, we stratified the Typically Developing control sample into three developmentally distinct age bins: Middle Childhood (5-10 years), Early Adolescence (10-14 years), and Late Adolescence/Young Adulthood (14-22 years). Subsequently, for each of the 23 interview questions, we calculated a bin-specific semantic baseline by computing the median embedding of answers provided by TD participants within that specific age range. We then computed the cosine similarity between each participant's response and the baseline of their corresponding age bin. This yielded a score representing how semantically 'typical' the response was. Participants with missing answers were assigned a missing score for that question. The interpretation of this score depends on the question's content; for narrative-recount questions, higher similarity reflects greater "typical narrative detail," whereas for questions about emotion, lower similarity may indicate more nuanced or idiosyncratic descriptions. Finally, these 23 individual typicality scores were averaged into conceptually meaningful domains based on the function of the questions. These domains included narrative detail, factual memory, thematic understanding, and perspective taking (e.g., describing one's own or a character's emotions, or the similarity between the dog and the kid, and how this relates to the kid's behavior). For example, the 'self-emotion description' score was the average semantic similarity score across answers to questions 14, 17, 20, and 23 (see Table S1 for details on these questions and examples of the 'Typical Answers' for each age bin). The Fisher-Z transform (inverse hyperbolic tangent, artanh) was applied to the semantic (cosine) similarity measures prior to regression analysis to approximate normal distributions.

Movement measures. We extracted anatomical and facial landmarks from video recordings at the native 30Hz sampling rate of the video recordings using Google's Mediapipe Holistic model (v.0.10.11) on a local machine. The raw 3D landmark coordinates were filtered to remove low-confidence data based on two criteria applied over a sliding window. First, segments with more than 10% missing values in a 2-second window were excluded. Second, segments exhibiting excessive jitter, defined as a standard deviation of the x-coordinate exceeding empirically-defined thresholds (0.1 for face; 0.01 for pose and landmarks) in 10-frame windows, were also excluded. A landmark point was considered valid for analysis only if it passed both of these checks. To ensure comparability across participants, the filtered data were normalized. Facial landmarks were aligned to a canonical model using an affine transformation[96] to correct for head rotation, translation, and scale. Body landmarks were normalized by the frame-by-frame inter-shoulder distance to render movements invariant to body size and participant's distance from the camera. Finally, movement was quantified as the frame-to-frame 3D Euclidean displacement for each landmark. These displacement values were then averaged within predefined anatomical regions (e.g., face, mouth, eyes, and body, see Fig. S7) and across all valid frames to yield a single mean movement score per region for each participant. Log of the average movement was used to model the effects of diagnosis on such measures.

Analytical Approach & Statistical Modeling

To test our primary research questions, we modeled the relationship between the measurements of interest and clinical diagnoses using multivariate linear regression. The model isolated the unique contribution of ASD and ADHD diagnostic status while controlling for key developmental and cognitive confounders. For each behavioral metric, the statistical model took the following form:

$$\text{Metric} \sim \text{Age} + \text{IQ} + \text{Sex} + \text{ADHD-Inatt} + \text{ADHD-Hyper} + \text{ASD} \quad (1)$$

with one model for each dependent variable (e.g., speech rate, lexical diversity), and the predictors were age (in years), a composite IQ score (Full-Scale), Sex (coded binary female=1, male=0), and binary variables for clinician-confirmed diagnoses of ASD, ADHD-Inattentive, and ADHD-Hyperactive (ADHD-Combined presentation was modeled as positive in both).
Every participant counts as n=1, with a set of numbers describing them: Age, IQ, ASD, ADHD-Hyper, ADHD-Inatt. For instance, an 8-year-old boy with an IQ of 90 who presents with ASD and ADHD-Combined is coded as: 8, 90, 0, 1, 1, 1. Models were fitted using the bisquare 'robust' linear fitting option in MATLAB (2024b). Variance Inflation Factors (VIF) confirmed that multicollinearity was not a confounding factor (all VIFs < 1.32). To quantify and compare the effect of each predictor, we estimated its effect size from the t statistic as follows:

Continuous variables: $$ES = t\sqrt{1/n} = f \quad (2)$$

Binary variables: $$ES = t\sqrt{1/n_1 + 1/n_2} = d \quad (3)$$

Here n is the degrees of freedom, and n_1 and n_2 are the number of data points for each binary group. These effect size measures allow for comparison across different variable types, even when the number of observations varies due to missing data for a specific behavioral metric. For the binary variables (Sex and Disorders), equation (3) is the traditional Cohen's d, and for continuous variables (Age and IQ), equation (2) captures the traditional Cohen's f. In both instances, it is the effect on the outcome variable over the residual unexplained error. Statistical tables for each model of the behavioral outcome measures, including all beta coefficients, standard errors, t-statistics, effect sizes, and p-values, are provided in Supplementary Sections S1.2-S1.5. All multivariate model p-values reported in the main manuscript text and figures are corrected for multiple comparisons using Bonferroni correction by multiplying model p-values by the number of tested behavioral features.

To assess diagnostic co-occurrence, we constructed three separate logistic regression models. For each model, one clinical diagnosis served as the binary outcome variable, while the remaining diagnoses were included as predictors, alongside age and sex. Crucially, the specific diagnosis used as the outcome was excluded from the predictor set for that model. We report odds ratios (OR) as the exponents of the beta coefficients (Fig. 2A and black arrows in Fig. 3C, Supplement Section S1.1):

$$\text{OR} = \exp(\beta). \quad (4)$$

Alternative analysis approaches

Our primary and planned modeling approach was designed to maximize the utility of this large, heterogeneous sample, capturing the full spectrum of age and diagnostic complexity. To validate these methodological decisions, we conducted several post-hoc sensitivity analyses:

First, although our models accounted for the significant overlap between ADHD and ASD, other co-occurring conditions could have influenced the results. However, we ruled out the impact of child anxiety on our behavioral measures by repeating all analyses while modeling for co-occurring positive anxiety diagnoses, and the associations with ADHD or ASD remained consistent (Fig. S10).

Second, although our reference group included individuals with other non-target DSM-5 diagnoses (N=653), excluding these participants to allow for a direct comparison with Typically Developing youths did not alter the significant associations found for ADHD or ASD.

Third, to mitigate potential bias from the skewed age distribution, we repeated the analysis, limiting it to the 5–14-year-old subset. All results remained the same, except that vocal loudness was no longer associated with ASD and ADHD-Hyperactive status. To clearly distinguish the variance related to Hyperactive versus Inattentive status, we did not separately code for a Combined presentation, unlike some prior studies[97–99]. When we repeated the analysis by coding with the Combined presentation instead of Hyperactive status, results on movement remained the same (Fig. S11).

Fourth, our planned analysis treated ASD, ADHD-Hyperactive, and ADHD-Inattentive as independent additive factors. This modeling choice is supported by the clinical instrument data in Table 1, which shows that the only significant trait difference between ASD with co-occurring ADHD (ASD[ADHD+]) and without (ASD[ADHD-]) lies in hyperactive and inattentive traits, whereas the autistic traits and other clinical measures remain comparable. We nevertheless explored whether co-occurring ADHD and ASD represent a distinct phenotypic category rather than an additive condition. When we modeled the data using mutually exclusive groups (ADHD-only, ASD-only, and ADHD+ASD), as is often done in the literature [49–51] the distinct separation of effects observed in our primary additive model was lost (Fig. S12), supporting our approach of treating these as independent, overlapping dimensions.

Finally, the linear model (1) does not include interaction terms. For age in particular, this means that the effects of the disorders are modeled as a constant gap between a positive and a negative diagnosis. However, for some behavioral features, the gap appears to increase with age (e.g., Fig. 3D, 4B). Such an overall gap can be detected by model (1) even if it increases with age. In a post-hoc analysis, we included the interaction of age with all diagnostic variables. None of the interactions were statistically significant after multiple comparison correction, while the Bayesian Information Criterion increased over the reduced model for most of the measures, suggesting that we are lacking statistical power to warrant the increase in model complexity. Narrowing it to variables with a main age and diagnosis effect suggests that there may be an

increasing gap with age for narrative detail in ASD (Fig. 3D, p=0.034 uncorrected), and for body movement in ADHD-Hyperactivity (Fig. 5C, p=0.044 uncorrected). Conversely, differences appear to diminish with age for pitch, pitch variation, and breathiness in ASD (p=0.004, 0.022 & 0.029, uncorrected). However, note that these are uncorrected and post-hoc analyses.

Visualization of Distinctive Terms

Semantic embedding analysis of the answers is quantitative and applies to all questions asked. To gain a sense of what it captures, we conducted post-hoc analyses on the significant ASD-question associations. To identify and visualize terms that distinguish ASD from non-ASD participant answers, we employed the *scattertext* library [57]. The horizontal axis represents the log frequency of each term in the corpus, while the vertical axis represents the term's association with the ASD group, calculated using a Scaled F-Score (β=2) and normalized via the normal cumulative distribution function to a range of [-1, 1], as done previously for the comparison of suicidal vs non-suicidal terms[100]. This scoring metric was selected to prioritize recall, where values approaching +1 indicate broad usage within the ASD group, and values approaching -1 indicate broad usage within the non-ASD control group. See Fig. S2 for the visualization of answers with a significant ASD association.

Ethics Statement

The data used in this study were collected between 2016 and 2022. The HBN protocol was approved by the Chesapeake Institutional Review Board. Written informed consent was obtained from participants aged 18 or older, and from legal guardians for participants younger than 18. Access to the data was governed by a Data Transfer and Use Agreement (dated 6/27/2024) between the Child Mind Institute and The City University of New York.

Data Availability

Phenotypical and diagnostic data are available upon request and with the establishment of a Data Use Agreement with the Child Mind Institute at **https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/.** Interview videos are not currently available due to the privacy of the children.

Code Availability

The full code pipeline used for speech transcription and diarization, as well as to extract language, vocal prosody, semantic embeddings, and body and face landmarks, is publicly available at **https://github.com/asortubay/AI_behavioral_analysis**.

Acknowledgments & Funding

NIH-NIMH P50 MH109429, CA Department of Health, grant to Nathan Kline Institute, Milham (PI). Gemini 3.1 Pro was used to assist in shortening our original text in the Introduction, Results and Discussion.

References

1. Kogan, M. D. *et al.* Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics* **124**, 1395–1403 (2009).
2. Polanczyk, G., De Lima, M. S., Horta, B. L., Biederman, J. & Rohde, L. A. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am. J. Psychiatry* **164**, 942–948 (2007).
3. Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C. & Rohde, L. A. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *Int. J. Epidemiol.* **43**, 434–442 (2014).
4. Yang, Y. *et al.* Prevalence of neurodevelopmental disorders among US children and adolescents in 2019 and 2020. *Front. Psychol.* **13**, 997648 (2022).
5. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. (American psychiatric association, 2013).
6. Grzadzinski, R. *et al.* Examining Autistic Traits in Children with ADHD: Does the Autism Spectrum Extend to ADHD? *J. Autism Dev. Disord.* **41**, 1178–1191 (2011).
7. Hattori, J. *et al.* Are pervasive developmental disorders and attention-deficit/hyperactivity disorder distinct disorders? *Brain Dev.* **28**, 371–374 (2006).
8. Hollingdale, J., Woodhouse, E., Young, S., Fridman, A. & Mandy, W. Autistic spectrum disorder symptoms in children and adolescents with attention-deficit/hyperactivity disorder: a meta-analytical review. *Psychol. Med.* **50**, 2240–2253 (2020).
9. Holtmann, M., Bölte, S. & Poustka, F. Attention deficit hyperactivity disorder symptoms in pervasive developmental disorders: association with autistic behavior domains and coexisting psychopathology. *Psychopathology* **40**, 172–177 (2007).
10. Nijmeijer, J. S. *et al.* Attention-deficit/hyperactivity disorder and social dysfunctioning. *Clin. Psychol. Rev.* **28**, 692–708 (2008).

11. Reiersen, A. M., Constantino, J. N., Volk, H. E. & Todd, R. D. Autistic traits in a population-based ADHD twin sample. *J. Child Psychol. Psychiatry* **48**, 464–472 (2007).
12. Simonoff, E. *et al.* Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J. Am. Acad. Child Adolesc. Psychiatry* **47**, 921–929 (2008).
13. Mulligan, A. *et al.* Autism symptoms in Attention-Deficit/Hyperactivity Disorder: A Familial trait which Correlates with Conduct, Oppositional Defiant, Language and Motor Disorders. *J. Autism Dev. Disord.* **39**, 197–209 (2009).
14. Reiersen, A. M., Constantino, J. N., Grimmer, M., Martin, N. G. & Todd, R. D. Evidence for shared genetic influences on self-reported ADHD and autistic symptoms in young adult Australian twins. *Twin Res. Hum. Genet.* **11**, 579–585 (2008).
15. Rommelse, N. N., Franke, B., Geurts, H. M., Hartman, C. A. & Buitelaar, J. K. Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *Eur. Child Adolesc. Psychiatry* **19**, 281–295 (2010).
16. Ronald, A., Simonoff, E., Kuntsi, J., Asherson, P. & Plomin, R. Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin sample. *J. Child Psychol. Psychiatry* **49**, 535–542 (2008).
17. Aoki, Y. *et al.* Association of White Matter Structure With Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry* **74**, 1120–1128 (2017).
18. Brieber, S. *et al.* Structural brain abnormalities in adolescents with autism spectrum disorder and patients with attention deficit/hyperactivity disorder. *J. Child Psychol. Psychiatry* **48**, 1251–1258 (2007).
19. Christakou, A. *et al.* Disorder-specific functional abnormalities during sustained attention in youth with attention deficit hyperactivity disorder (ADHD) and with autism. *Mol. Psychiatry* **18**, 236–244 (2013).

20. Di Martino, A. *et al.* Shared and Distinct Intrinsic Functional Network Centrality in Autism and Attention-Deficit/Hyperactivity Disorder. *Biol. Psychiatry* **74**, 623–632 (2013).
21. Bedford, S. A. *et al.* Brain-Charting Autism and Attention-Deficit/Hyperactivity Disorder Reveals Distinct and Overlapping Neurobiology. *Biol. Psychiatry* **97**, 517–530 (2025).
22. Norman, L. J. *et al.* Cross-sectional mega-analysis of resting-state alterations associated with autism and attention-deficit/hyperactivity disorder in children and adolescents. *Nat. Ment. Health* **3**, 709–723 (2025).
23. Alexander, L. M. *et al.* An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* **4**, 170181 (2017).
24. Geurts, H. M. & Embrechts, M. Language Profiles in ASD, SLI, and ADHD. *J. Autism Dev. Disord.* **38**, 1931–1943 (2008).
25. Bishop, D. V. & Baird, G. Parent and teacher report of pragmatic aspects of communication: use of the Children's Communication Checklist in a clinical setting. *Dev. Med. Child Neurol.* **43**, 809–818 (2001).
26. Geurts, H. M. *et al.* Can the Children's Communication Checklist differentiate between children with autism, children with ADHD, and normal controls? *J. Child Psychol. Psychiatry* **45**, 1437–1453 (2004).
27. Geurts, H. M. & Embrechts, M. Language profiles in ASD, SLI, and ADHD. *J. Autism Dev. Disord.* **38**, 1931–1943 (2008).
28. Kim, Y.-S. G., Dore, R., Cho, M., Golinkoff, R. & Amendum, S. J. Theory of mind, mental state talk, and discourse comprehension: Theory of mind process is more important for narrative comprehension than for informational text comprehension. *J. Exp. Child Psychol.* **209**, 105181 (2021).
29. Baron-Cohen, S., Leslie, A. M. & Frith, U. Does the autistic child have a "theory of mind"? *Cognition* **21**, 37–46 (1985).

30. Baron-Cohen, S., Leslie, A. M. & Frith, U. Does the autistic child have a “theory of mind” ? *Cognition* **21**, 37–46 (1985).
31. Tager-Flusberg, H., Paul, R. & Lord, C. Language and communication in autism. *Handb. Autism Pervasive Dev. Disord.* **1**, 335–364 (2005).
32. Kuijper, S. J. M., Hartman, C. A., Bogaerds-Hazenberg, S. T. M. & Hendriks, P. Narrative production in children with autism spectrum disorder (ASD) and children with attention-deficit/hyperactivity disorder (ADHD): Similarities and differences. *J. Abnorm. Psychol.* **126**, 63–75 (2017).
33. Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M. & Gaigg, S. B. “Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis”. *Autism Res.* **10**, 384–407 (2017).
34. Trayvick, J. *et al.* Speech and language patterns in autism: Towards natural language processing as a research and clinical tool. *Psychiatry Res.* **340**, 116109 (2024).
35. Hamdan, A.-L. *et al.* Vocal Characteristics in Children With Attention Deficit Hyperactivity Disorder. *J. Voice* **23**, 190–194 (2009).
36. Barona-Lleo, L. & Fernandez, S. Hyperfunctional Voice Disorder in Children With Attention Deficit Hyperactivity Disorder (ADHD). A Phenotypic Characteristic? *J. Voice* **30**, 114–119 (2016).
37. Briend, F. *et al.* Voice acoustics allow classifying autism spectrum disorder with high accuracy. *Transl. Psychiatry* **13**, 250 (2023).
38. Eni, M. *et al.* Reliably quantifying the severity of social symptoms in children with autism using ASDSpeech. *Transl. Psychiatry* **15**, 14 (2025).
39. Godel, M. *et al.* Prosodic signatures of ASD severity and developmental delay in preschoolers. *Npj Digit. Med.* **6**, 99 (2023).

40. de Belen, R. A. J., Bednarz, T., Sowmya, A. & Del Favero, D. Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. *Transl. Psychiatry* **10**, 1–20 (2020).
41. Zampella, C. J., Wang, L. A. L., Haley, M., Hutchinson, A. G. & de Marchena, A. Motor Skill Differences in Autism Spectrum Disorder: a Clinically Focused Review. *Curr. Psychiatry Rep.* **23**, 64 (2021).
42. De Crescenzo, F. *et al.* The use of actigraphy in the monitoring of sleep and activity in ADHD: A meta-analysis. *Sleep Med. Rev.* **26**, 9–20 (2016).
43. Hartman, C. A., Geurts, H. M., Franke, B., Buitelaar, J. K. & Rommelse, N. N. J. Changing ASD-ADHD symptom co-occurrence across the lifespan with adolescence as crucial time window: Illustrating the need to go beyond childhood. *Neurosci. Biobehav. Rev.* **71**, 529–541 (2016).
44. Larsson, H., Dilshad, R., Lichtenstein, P. & Barker, E. D. Developmental trajectories of DSM-IV symptoms of attention-deficit/hyperactivity disorder: genetic effects, family risk and associated psychopathology. *J. Child Psychol. Psychiatry* **52**, 954–963 (2011).
45. Loomes, R., Hull, L. & Mandy, W. P. L. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 466–474 (2017).
46. Ramtekkar, U. P., Reiersen, A. M., Todorov, A. A. & Todd, R. D. Sex and age differences in Attention-Deficit/Hyperactivity Disorder symptoms and diagnoses: Implications for DSM-V and ICD-11. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 217-28.e1–3 (2010).
47. Joseph, R. M. & Fein, D. The significance of IQ and differential cognitive abilities for understanding ASD. *Neuropsychol. Autism* **281**, 294 (2011).
48. Charman, T. *et al.* Defining the cognitive phenotype of autism. *Brain Res.* **1380**, 10–21 (2011).

49. Craig, F. *et al.* Overlap Between Autism Spectrum Disorders and Attention Deficit Hyperactivity Disorder: Searching for Distinctive/Common Clinical Features. *Autism Res.* **8**, 328–337 (2015).
50. Rao, P. A. & Landa, R. J. Association between severity of behavioral phenotype and comorbid attention deficit hyperactivity disorder symptoms in children with autism spectrum disorders. *Autism* **18**, 272–280 (2014).
51. Scandurra, V. *et al.* Neurodevelopmental Disorders and Adaptive Functions: A Study of Children With Autism Spectrum Disorders (ASD) and/or Attention Deficit and Hyperactivity Disorder (ADHD). *Front. Psychiatry* **10**, (2019).
52. Simonoff, E. *et al.* Psychiatric Disorders in Children With Autism Spectrum Disorders: Prevalence, Comorbidity, and Associated Factors in a Population-Derived Sample. *J. Am. Acad. Child Adolesc. Psychiatry* **47**, 921–929 (2008).
53. Menyuk, P. & Brisk, M. E. Language Development in Middle Childhood: Ages 9–13. in *Language development and education: Children with varying language experiences* 118–141 (Springer, 2005).
54. Wechsler, D. *Wechsler Intelligence Scale for Children-Fifth Edition*. (Pearson, Bloomington, MN).
55. Wechsler, D. Wechsler abbreviated scale of intelligence--. (1999).
56. Kaufman, A. S. & Kaufman, N. L. Kaufman brief intelligence test. *Encycl. Spec. Educ. Ref. Educ. Child. Adolesc. Adults Disabil. Except. Individ.* (2013).
57. Kessler, J. Scattertext: a Browser-Based Tool for Visualizing how Corpora Differ. in *Proceedings of ACL 2017, System Demonstrations* (eds Bansal, M. & Ji, H.) 85–90 (Association for Computational Linguistics, Vancouver, Canada, 2017).
58. Worthington, M., Efstathiadis, G., Yadav, V. & Abbas, A. 172. OpenWillis: An Open-Source Python Library for Digital Health Measurement. *Biol. Psychiatry* **95**, S169–S170 (2024).

59. Korrel, H., Mueller, K. L., Silk, T., Anderson, V. & Sciberras, E. Research Review: Language problems in children with Attention-Deficit Hyperactivity Disorder–a systematic meta-analytic review. *J. Child Psychol. Psychiatry* **58**, 640–654 (2017).
60. Bellani, M., Moretti, A., Perlini, C. & Brambilla, P. Language disturbances in ADHD. *Epidemiol. Psychiatr. Sci.* **20**, 311–315 (2011).
61. Sciberras, E. *et al.* Language problems in children with ADHD: A community-based study. *Pediatrics* **133**, 793–800 (2014).
62. Tirosh, E. & Cohen, A. Language deficit with attention-deficit disorder: A prevalent comorbidity. *J. Child Neurol.* **13**, 493–497 (1998).
63. Redmond, S. M. Conversational profiles of children with ADHD, SLI and typical development. *Clin. Linguist. Phon.* **18**, 107–125 (2004).
64. Jepsen, I. B., Hougaard, E., Matthiesen, S. T. & Lambek, R. A systematic review and meta-analysis of narrative language abilities in children with Attention-Deficit/Hyperactivity Disorder. *Res. Child Adolesc. Psychopathol.* **50**, 737–751 (2022).
65. Kuijper, S. J., Hartman, C. A., Bogaerds-Hazenberg, S. & Hendriks, P. Narrative production in children with autism spectrum disorder (ASD) and children with attention-deficit/hyperactivity disorder (ADHD): Similarities and differences. *J. Abnorm. Psychol.* **126**, 63 (2017).
66. Geurts, H. M., Broeders, M. & Nieuwland, M. S. Thinking outside the executive functions box: Theory of mind and pragmatic abilities in attention deficit/hyperactivity disorder. *Eur. J. Dev. Psychol.* **7**, 135–151 (2010).
67. Tannock, R. & Schachar, R. Executive dysfunction as an underlying mechanism of behavior and language problems in attention deficit hyperactivity disorder. in *Language, learning, and behavior disorders: Developmental, biological, and clinical perspectives* 128–155 (Cambridge University Press, New York, NY, US, 1996).

68. Taurines, R. *et al.* ADHD and autism: differential diagnosis or overlapping traits? A selective review. *ADHD Atten. Deficit Hyperact. Disord.* **4**, 115–139 (2012).

69. Naigles, L. R. *et al.* “You’re telling me!” The prevalence and predictors of pronoun reversals in children with autism spectrum disorders and typical development. *Res. Autism Spectr. Disord.* **27**, 11–20 (2016).

70. Diehl, J. J., Bennetto, L. & Young, E. C. Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *J. Abnorm. Child Psychol.* **34**, 83–98 (2006).

71. Baron-Cohen, S., Leslie, A. M. & Frith, U. Does the autistic child have a “theory of mind” ? *Cognition* **21**, 37–46 (1985).

72. Losh, M. & Gordon, P. C. Quantifying Narrative Ability in Autism Spectrum Disorder: A Computational Linguistic Analysis of Narrative Coherence. *J. Autism Dev. Disord.* **44**, 3016–3025 (2014).

73. Lee, M. *et al.* What’s the story? A computational analysis of narrative competence in autism. *Autism* **22**, 335–344 (2018).

74. Fusaroli, R. *et al.* Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children. *Autism Res.* **15**, 653–664 (2022).

75. Pecoraro, P. M., Marsili, L., Espay, A. J., Bologna, M. & di Biase, L. Computer Vision Technologies in Movement Disorders: A Systematic Review. *Mov. Disord. Clin. Pract.* (2025).

76. Cook, A., Mandal, B., Berry, D. & Johnson, M. Towards Automatic Screening of Typical and Atypical Behaviors in Children With Autism. in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* 504–510 (2019). doi:10.1109/DSAA.2019.00065.

77. Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T. & Schaer, M. Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Sci. Rep.* **11**, 15069 (2021).
78. Lemler, C. *et al.* Semi-Automated Multi-Label Classification of Autistic Mannerisms by Machine Learning on Post Hoc Skeletal Tracking. *Autism Res.* **18**, 833–844 (2025).
79. Vyas, K. *et al.* Recognition Of Atypical Behavior In Autism Diagnosis From Video Using Pose Estimation Over Time. in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)* 1–6 (2019). doi:10.1109/MLSP.2019.8918863.
80. Zhang, Y., Tian, Y., Wu, P. & Chen, D. Application of Skeleton Data and Long Short-Term Memory in Action Recognition of Children with Autism Spectrum Disorder. *Sensors* **21**, 411 (2021).
81. Koehler, J. C. *et al.* Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions. *Transl. Psychiatry* **14**, 76 (2024).
82. Faerman, A. *et al.* Language Abilities are Associated with Both Verbal and Nonverbal Intelligence in Children on the Autism Spectrum. *Dev. Neuropsychol.* **48**, 248–257 (2023).
83. Dennis, M. *et al.* Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *J. Int. Neuropsychol. Soc.* **15**, 331–343 (2009).
84. Lord, C. *et al.* The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30**, 205–223 (2000).
85. Stevens, T., Peng, L. & Barnard-Brak, L. The comorbidity of ADHD in children diagnosed with autism spectrum disorder. *Res. Autism Spectr. Disord.* **31**, 11–18 (2016).
86. Zablotsky, B., Bramlett, M. D. & Blumberg, S. J. The Co-Occurrence of Autism Spectrum Disorder in Children With ADHD. *J. Atten. Disord.* **24**, 94–103 (2020).

87. Kaufman, J. *et al.* Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J. Am. Acad. Child Adolesc. Psychiatry* **36**, 980–988 (1997).

88. Swanson, J. M. *et al.* Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. *Int. J. Educ. Psychol. Assess.* **10**, 51–70 (2012).

89. Ehlers, S., Gillberg, C. & Wing, L. A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *J. Autism Dev. Disord.* **29**, 129–141 (1999).

90. Angold, A., Costello, E. J., Messer, S. C. & Pickles, A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int. J. Methods Psychiatr. Res.* (1995).

91. Birmaher, B. *et al.* Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. *J. Am. Acad. Child Adolesc. Psychiatry* **38**, 1230–1236 (1999).

92. Achenbach, T. M. *Integrative Guide for the 1991 CBCL/4-18, YSR, and TRF Profiles*. (Department of Psychiatry, University of Vermont, 1991).

93. Bain, M., Huh, J., Han, T. & Zisserman, A. WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. in *INTERSPEECH 2023* 4489–4493 (ISCA, 2023). doi:10.21437/Interspeech.2023-78.

94. Bredin, H. *et al.* Pyannote. audio: neural building blocks for speaker diarization. in 7124–7128 (IEEE, 2020).

95. Lee, J. *et al.* Gecko: Versatile Text Embeddings Distilled from Large Language Models. Preprint at https://doi.org/10.48550/arXiv.2403.20327 (2024).

96. Silvan, A., Parra, L. C. & Madsen, J. Real-Time Estimation of Overt Attention from Dynamic Features of the Face Using Deep Learning. in *2025 47th Annual International Conference of*

the IEEE Engineering in Medicine and Biology Society (EMBC) 1–5 (2025). doi:10.1109/EMBC58623.2025.11254013.

97. Conway, K. P. *et al.* Association between COVID-19 risk-mitigation behaviors and specific mental disorders in youth. *Child Adolesc. Psychiatry Ment. Health* **17**, 14 (2023).
98. Restrepo, A. *et al.* Problematic internet use in children and adolescents: associations with psychiatric disorders and impairment. *BMC Psychiatry* **20**, 252 (2020).
99. Kramer, E. *et al.* Diagnostic Associations of Processing Speed in a Transdiagnostic, Pediatric Sample. *Sci. Rep.* **10**, 10114 (2020).
100. Zuromski, K. L. *et al.* Detecting suicide risk among U.S. servicemembers and veterans: a deep learning approach using social media data. *Psychol. Med.* **54**, 3379–3388 (2024).

Supplementary Material

Table S1. Interview Questions & Typical Answers 35
Table S2. Detailed Statistical Test Results for Table 1 37
Table S3. Description of Behavioral Variables. 38
Figure S1. Co-ocurring diagnoses and age distributions. 42
Figure S2. Scattertext visualization of used terms when answering specific questions. 43
Figure S3. Results are independent of including IQ as covariate. 44
Figure S4. Comparing Clinical Instruments to measured behavior. 45
Figure S5. Unique association of clinical questionnaire scales with clinician-confirmed diagnoses. 46
Figure S6. Distribution of Autism Scores in the ASD-positive population. 46
Figure S7. Interview video examples, face landmarks, and validity. 47
Figure S8. Validity of interview questions. 48
Figure S9. Distribution of Behavioral Variables. 49
Figure S10. Results are independent of anxiety diagnoses. 50
Figure S11. Results are independent of how the ADHD presentations are coded. 51
Figure S12. Modeling ASD(ADHD+) and ASD(ADHD-) separately does not paint a clear picture. 52
Supplementary Section S1. Multivariate Model Reports. 52
- Supplementary Section S1.1. Diagnoses 52
- Supplementary Section S1.2. Language Use Measures 53
- Supplementary Section S1.3. Semantic Measures 57
- Supplementary Section S1.4. Vocal Prosody Measures 60
- Supplementary Section S1.5. Movement Measures 62

Supplementary Section S2. Diarizing Prompt for LLM 64
Supplementary Section S3. Question-Answer Extraction Prompt for LLM 65

Table S1. Interview Questions & Typical Answers

Semantic Measure	No.	Question	Typical Answers*		
			4-10 years	10-14 years	14-22 years
—	1	"I hope you enjoyed the last movie. Have you seen it before?"	No.	No?	Yeah.
Narrative Detail	2	"Can you tell me what happened in the movie? Try to tell the whole story. Remember that stories have a beginning, things that happen, and an ending."	Well, so he was playing a video game. And he was just like, mom. He was playing a video game, and he was sucked into it. And he was like, mm. And then his mom came in the room and, well, came back. from the house and said, and I forgot what she said, but she put a box in front of him and he was like, what is this? Oh my God. And he kept on playing. The mother said, why don't you look what I got for you in the box? And then she got a phone call and the kid was like, for me? And then he opened it and then he saw the puppy and he was very grateful, but then he saw the puppy had no paw and he was like, and he threw him away. And the puppy was just like, and then the puppy went back to him and he just threw him away again. Then the puppy just found like this red ball and the puppy was playing with the ball because it was his first time meeting a ball, I think. And the ball went to the went to the boy and the puppy was like throw the ball for me please and the boy just kicked the ball because he was angry just because yeah and the puppy was playful and kept on running and fell and run fell and run	Well, the boy was playing video games and his mom came in with a box and she said it was a present for him. So he looked at it and he didn't really care and he was still involved in his video game. And then the mom took a phone call and he noticed that she left. So he opened the box to see what was in it. And it was a puppy and it came out. And then one of his legs was missing. So the boy threw him away and didn't want him. And then the dog found a ball and started playing with it. The boy, the dog tried to get the boy to play with him. And then after a while, the boy like gave in and started playing with him. And the boy had one leg too, so. And then they went outside and played fetch, I think.	OK. So it started off with the boy. He seemed kind of upset, and he was just playing video games. And it was kind of dark in their house and stuff. And his mom comes home, and she like turns on the lights and everything and she gives him, she gives him the gift in the box so he stops playing and then he opens the box and he like sees a puppy and then he kind of gets excited and then he sees that the puppy doesn't have a leg, like is missing one leg. So then he's like not interested and he kind of just puts it off to the side but the dog doesn't really know what's going on. It's just like excited to be with someone like an owner or something. And then he goes back to playing video games, and the dog is kind of just searching around the house and finds the ball. And you can tell that the dog's kind of having trouble walking and playing with it and stuff. And so then he just kind of watches the owner. He kind of watches as the dog's kind of struggling. but he's playing with the ball and it's kind of making him kind of happy watching him just like play around and he's like still happy. And then he eventually turns off his video game and then it ends when they go outside to play. And you realize that he has some, like he, his leg also is missing.

			and um The puppy went back to the boy, and the boy, well, no, no, no. And then the boy started smiling, and then he was like, oh, no, no. And then... He kept on smiling, and he was like, okay. And then the puppy went then back to the boy, and the boy got up from his seat, and he didn't have a leg. The same as with the puppy.		
—	3	"Do you remember anything else from the story?"	No.	No?	Um, no. Okay.
Positive Preferences	4	"What are some of the things you liked about the movie?"	I liked the dog. I liked the dog. He was really, really cute.	I liked how I liked how, like, the dog and, um, the boy, like, like, um, how they got along and stuff like that at the end.	I liked the dog. It was cute. I can't think of what else I liked. I liked the overall story of it.
Negative Preferences	5	"Is there something you didn’t like about the movie?"	That he kicked it. And he didn't like the dog.	There wasn't a whole lot that I didn't like. I mean, like, honestly, I didn't like the fact that he was being mean in the beginning, kicking the dog.	Well, I kind of felt upset about the dog being pushed around at first, but other than that, everything else was fine. Yeah.
Factual Memory	6	"Who gave the boy a box?"	His mom.	His mom.	His mom.
Factual Memory	7	"What was in the box?"	A puppy.	A puppy.	A dog.
Factual Memory	8	"What was the boy doing before he got the box?"	Playing video games.	Playing video games.	Playing video games.
Factual Memory	9	"What was the puppy playing with?"	A ball.	A ball.	A ball. Red ball.
Thematic Understanding	10	"How are the puppy and the boy the same?"	But they both have one missing leg.	They're both missing a leg.	They both are missing a leg.
—	11	"In the movie, who is missing a leg? The boy, the puppy, or no one?"	Both, the boy and the puppy.	Both.	Both.
Kids rewatch a short clip from "The Present" (00:53-01:02)					
External Emotion Description	12	"How do you think the	Happy.	Happy, excited.	Happy.

		puppy was feeling?"			
External Emotion Description	13	"How do you think the boy was feeling?"	Happy.	Excited, happy.	Happy, excited.
Self Emotion Description	14	"And how did you feel while you were watching that part?"	Happy.	Happy.	Happy.
Kids rewatch a short clip from "The Present" (01:03-01:09)					
External Emotion Description	15	"How do you think the puppy was feeling?"	Sad.	Um, like sad and disappointed.	At first he was okay and then he was probably upset.
External Emotion Description	16	"How do you think the boy was feeling?"	Um, mad.	Disgusted and mad.	Uh, upset, disappointed.
Self Emotion Description	17	"And how did you feel while you were watching that part?"	Sad.	I was feeling kind of mad. Yeah. Or just sad.	Kind of felt bad for the dog.
Kids rewatch a short clip from "The Present" (01:03-01:09)					
External Emotion Description	18	"How do you think the puppy was feeling?"	Sad.	Sad. Like he tried again to like play and all that and then he just kicked him over to get lost.	The puppy was still a little bit happy, but then once he got kicked away, pretty sure he felt disregarded. He felt bad. Yeah, he felt bad.
External Emotion Description	19	"How do you think the boy was feeling?"	Um, mad.	Um, angry.	Angry. Still mad.
Self Emotion Description	20	"And how did you feel while you were watching that part?"	Sad.	I felt kind of mad.	A little bit sadder because he was actually kind of unleashing his emotions at the puppy.
Kids rewatch a short clip from "The Present" (01:14-01:23)					
External Emotion Description	21	"How do you think the puppy was feeling?"	Happy.	Happy.	Happy.
External Emotion Description	22	"How do you think the boy was feeling?"	Happy.	Happy.	Happy, too.
Self Emotion Description	23	"And how did you feel while you were watching that part?"	Happy.	Um... Happy.	Yeah, happy.

* Typical Answers are those that have the highest cosine similarity to the median embedding of the age-stratified Typically Developing population.

Table S2. Detailed Statistical Test Results for Table 1

Variable	ADHD-Inatt vs TD	ADHD-Hyper vs TD	ASD(ADHD+) vs TD	ASD(ADHD-) vs TD	Other vs TD	ASD(ADHD+) vs ASD(ADHD-)
Age	Z=3.30, p=0.005	Z=-0.50, p=1.000	Z=2.45, p=0.072	Z=1.29, p=0.982	Z=3.93, p<0.001	Z=0.05, p=0.963

Sex (Distribution)	X2(1)=20.76, p<0.001	X2(1)=32.16, p<0.001	X2(1)=50.25, p<0.001	X2(1)=3.43, p=0.319	X2(1)=0.01, p=1.000	X2(1)=10.34, p=0.001
Inattention (SWAN)	Z=15.06, p<0.001	Z=13.69, p<0.001	Z=12.10, p<0.001	Z=4.96, p<0.001	Z=2.70, p=0.035	t(338.0)=4.45, p<0.001
Hyperactivity (SWAN)	Z=11.53, p<0.001	Z=15.85, p<0.001	Z=11.90, p<0.001	Z=4.28, p<0.001	Z=-0.16, p=1.000	t(338.0)=4.45, p<0.001
Autism (ASSQ)	Z=9.90, p<0.001	Z=11.29, p<0.001	Z=15.05, p<0.001	Z=10.36, p<0.001	Z=2.88, p=0.020	t(346.0)=-0.49, p=0.627
Depression (MFQ)	Z=8.06, p<0.001	Z=8.18, p<0.001	Z=7.39, p<0.001	Z=4.04, p<0.001	Z=4.14, p<0.001	Z=1.06, p=0.289
Anxiety (SCARED)	Z=5.12, p<0.001	Z=4.79, p<0.001	Z=5.56, p<0.001	Z=5.57, p<0.001	Z=4.90, p<0.001	Z=-1.95, p=0.051
Externalizing (CBCL)	Z=9.44, p<0.001	Z=12.43, p<0.001	Z=8.30, p<0.001	Z=4.16, p<0.001	Z=3.18, p=0.007	Z=1.49, p=0.135
Internalizing (CBCL)	Z=8.38, p<0.001	Z=8.28, p<0.001	Z=9.46, p<0.001	Z=6.45, p<0.001	Z=6.16, p<0.001	Z=-0.44, p=0.659
Full-Scale IQ	t(1361.0)=-5.91 , p<0.001	t(856.0)=-4.47, p<0.001	t(430.0)=-5.80, p<0.001	t(222.0)=-3.91, p<0.001	t(749.0)=-5.49, p<0.001	t(328.0)=-0.08, p=0.932
Verbal IQ	Z=-4.34, p<0.001	Z=-3.26, p=0.006	t(430.0)=-5.10, p<0.001	t(222.0)=-3.96, p<0.001	t(751.0)=-5.42, p<0.001	t(328.0)=0.37, p=0.715
Interview Length	Z=0.71, p=1.000	Z=-0.14, p=1.000	Z=1.71, p=0.434	Z=1.51, p=0.652	Z=0.76, p=1.000	Z=-0.54, p=0.587

Statistical differences for continuous measures were assessed using independent samples t-tests (for normally distributed data) or Wilcoxon rank-sum tests (for non-normally distributed data), following normality checks via the Kolmogorov-Smirnov test. Categorical comparisons were performed using Chi-square tests. * $p < 0.01$, ** $p < 0.001$ compared to the Typically Developing group (Bonferroni corrected for 5 comparisons). † $p < 0.01$, †† $p < 0.001$ for the comparison between ASD(ADHD+) and ASD(ADHD-).

Table S3. Description of Behavioral Variables.

Group	Measured Aspect	Variable	Variable Name *	Measurement Description	Extraction Method
Language Ability Measures	**Structural Language**	*Amount of Speech*	*speech_length_words*	Total subject speech length in words	Total number of words spoken by the subject in the transcribed & diarized transcript. Log (ln) of total words was used to approximate a normal distribution.
		Speech Rate	*words_per_min*	Subject speech rate, in words per minute	Total subject words divided by subject speech length in minutes
		Filler Word Usage		Number of filler words per minute of subject speech	Total no. of fillers ('um', 'uh', 'er', 'ah', 'like', 'you know', 'so', 'okay', 'right') in subject speech divided by subject speech length in minutes

		% of unique words	*unique_word_percentage*	Percentage of unique words in total subject speech	100 - repeated words percentage, measured as the percentage of words in each utterance that are repeated. This is calculated by using a sliding window of 10 words to adjust for utterance length.
		Lexical Diversity	*mattr_10*	Average Moving Average Type Token Ratio with a window size of 10	Moving Average Type Token Ratio (MATTR), computed using the average of TTRs (computed as t/w, where t is the number of unique terms/vocab, and w is the total number of words) over successive segments of a text by taking the estimate TTR for tokens 1 to n, 2 to n+1, 3 to n+2, and so on until the end of the text (where n is window size), then taking the average. Extracted here using the library *lexicalrichness* with a window size of 10 on the subject utterances.
		Between Word Coherence	*word_coherence_mean*	Average measure of the semantic similarity (cosine) of each word to the preceding word using LLMs	Average semantic similarity of each word to the immediately preceding word, computed by the cosine similarity between the BERT embeddings (*bert-base-cased*) of each word.
	Pragmatic Language	*Subject Conversation Presence*	*speech_percentage*	Time spent speaking by the subject over the total time	Sum of all the subject speaking times divided by the total conversation time.
		Self-Focused Speech	*first_person_percentage*	Percentage of pronouns in the subject speech that are first-person singular pronouns	Percentage of ("I", "me", "my", "mine", "myself") occurences in the subject speech divided by total pronoun occurences, identified by *nltk*'s PRP (Personal Pronouns) and PRP$ (Possesive Pronouns).

		Question-Answer Coherence	*turn_to_turn_tangeniality_mean*	Average semantic similarity of the current speaker's turn to the previous turn of the other speaker	Average semantic similarity of each subject turn to the immediately preceding interviewer turn, computed by the cosine similarity between the *sentence-transformers/all-MiniLM-L6-v2* text embeddings of each turn.
Semantic Measures (Socio-Cognitive Abilities)	**Memory**	*Factual Memory*		Answer Typicality when asked about facts in the movie	Semantic similarity (cosine similarity) of each participant answer's 768-dimensional text embedding (Google's *text-embedding-004*), with the corresponding age-matched median 'typical' answer embedding. The Fisher-Z transform (inverse hyperbolic tangent, atanh) was applied to the semantic (cosine) similarity measures prior to regression analysis to approximate normal distributions.
	Social Judgment & Salience	*External Emotion Description*		Answer Typicality when asked about dog/child emotions in the movie clips	
	Theory of Mind	*Self Emotion Description*		Answer Typicality when asked about self-emotions in the movie clips	
	Narrative Production, Comprehension & Theory of Mind	*Narrative Detail*		Answer Typicality when asked to give a detailed recall of the movie	
	Theory of Mind	*Thematic Understanding*		Answer Typicality when asked about the similarities between the dog and the kid, and their shared experience	
	Social Judgment & Salience	*Positive Preferences*		Answer Typicality when asked about what they liked in the movie	
	Social Judgment & Salience	*Negative Preferences*		Answer Typicality when asked about what they did not like in the movie	
Vocal Prosody Measures		*Pitch*	*f0_mean*	Vocal Pitch (f0, Hz)	Average and standard deviation for the framewise subject fundamental frequency (f0), extracted in the 75-500 Hz range using *Parselmouth*'s *praat* tool.
		Pitch Variation	*f0_stddev*	Pitch variation (std(f0))	
		Loudness	*loudness_mean*	Vocal intensity in dB	Average and standard deviation for the framewise subject

		Loudness Variation	*loudness_stddev*	Intensity variation (std(dB))	loudness (in dB), extracted using *Parselmouth*'s *praat* tool.
		Dysphonia	*cpp_mean*	Mean cepstral peak prominence	Average framewise subject Cepstral Peak Prominence (Fraile & Godino-Llorente, 2014).
		Breathiness	*gne_ratio*	Glottal-to-noise excitation ratio	Glottal-to-noise excitation ratio (Godino-Llorente et al, 2010), extracted using *Parselmouth*'s tool.
Movement Measures		*Facial Movements*		Total movement in the video, per region.	Anatomical and facial landmarks were extracted using Google's Mediapipe Holistic model (model_complexity=2). Then filtered to remove low-confidence or jittery segments and normalized to correct for head orientation and body size. Movement was quantified as the frame-to-frame 3D Euclidean displacement, averaged within specific anatomical regions for each participant. Finally, these regional averages were log-transformed to approximate a normal distribution.
		Eye Movements			
		Mouth Movements			
		Body Movements			

*Variable Names for the specific measures that were extracted using the OpenWillis tools. For Language Ability Mesures, using speech_characteristics function, with the optional settings "min_turn_length = 5, min_coherence_turn_length = 5, option = 'coherence'". For Vocal Prosody Measures, using the vocal_acoustics function, with the optional "voiced_segments = True" setting to onlly compute measures in voice segments longer than 100 ms.

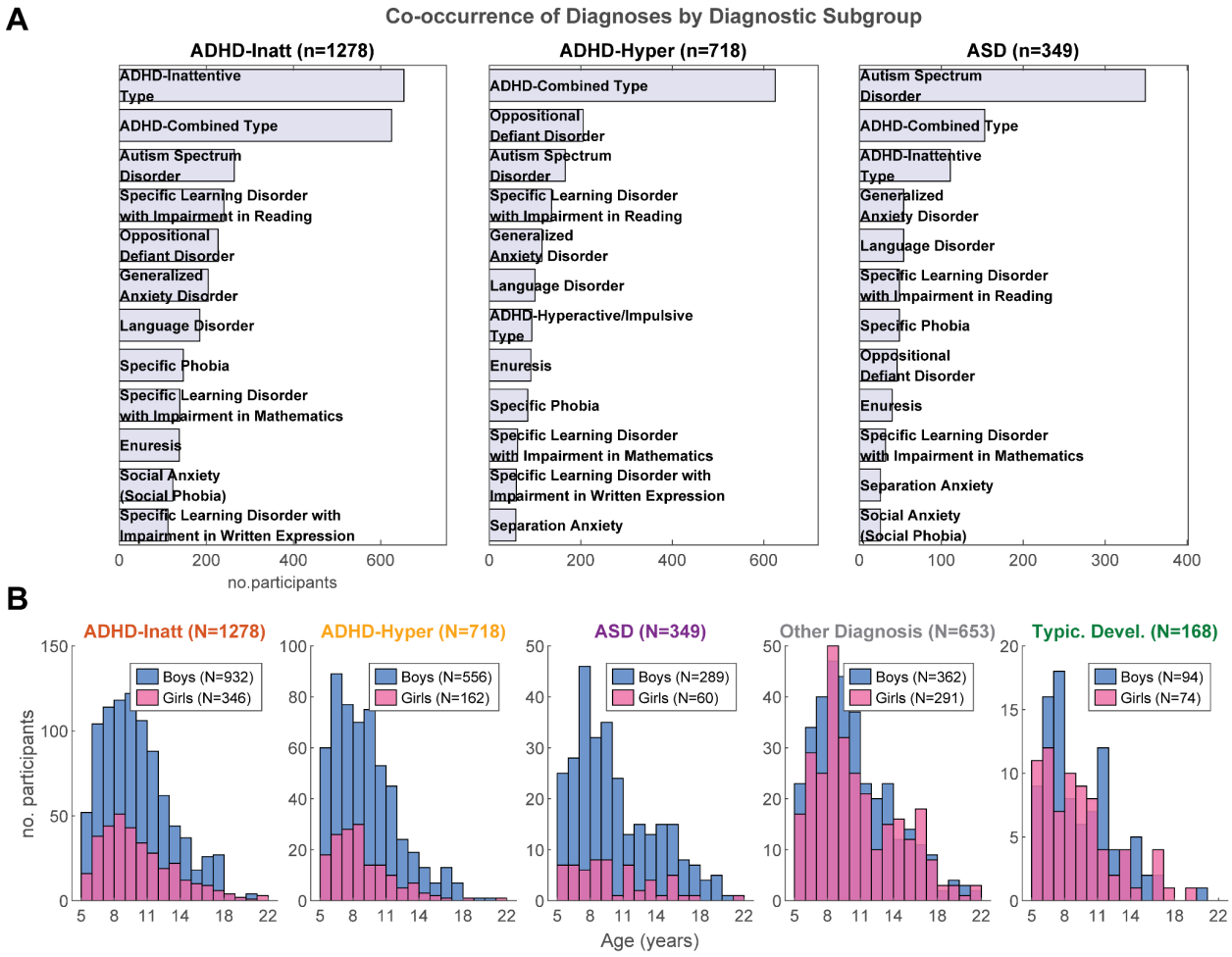

Figure S1. Co-occurring diagnoses and age distributions.

A) Co-occurring diagnoses (Top 12) in each of the three diagnostic groups of interest. B) Age-Sex distribution across the diagnostic groups. ADHD-Inatt and ADHD-Hyper contain participants with an ADHD-Combined presentation. Other Diagnosis grop contains participants with some positive diagnosis, excluding those with ADHD or ASD.

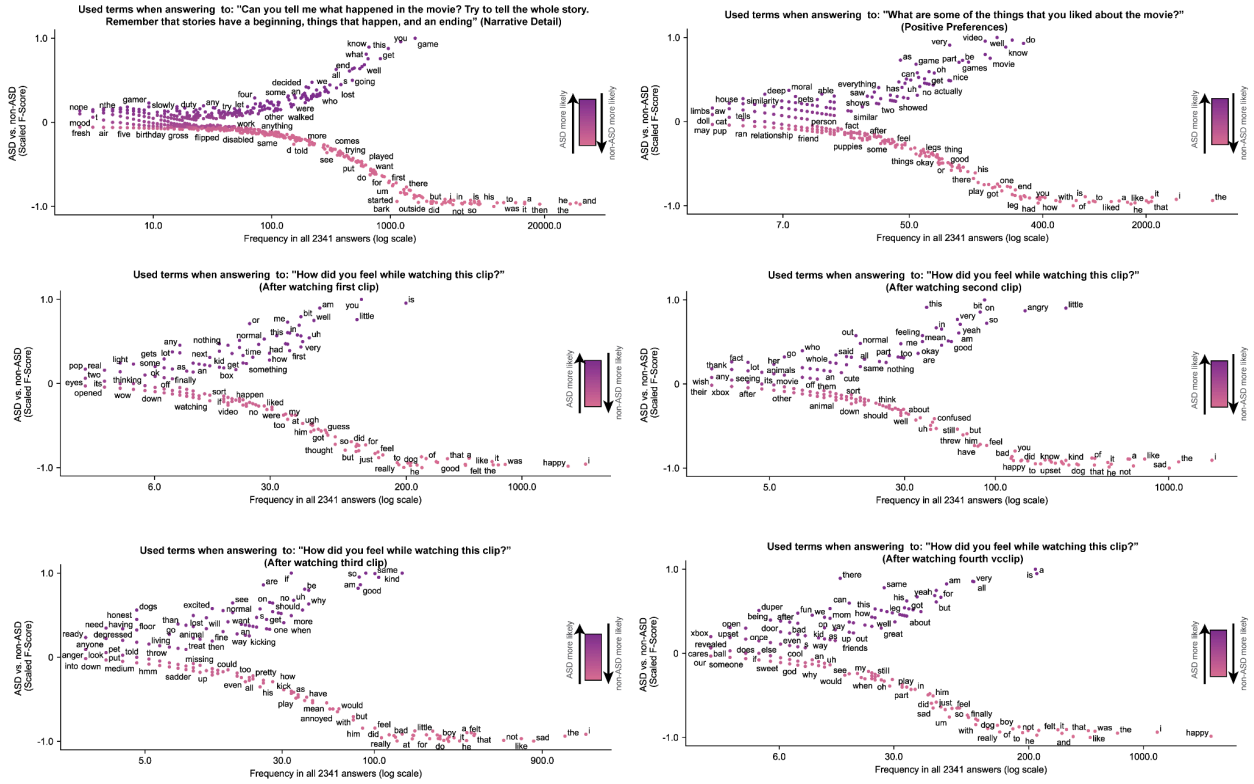

Figure S2. Scattertext visualization of used terms when answering specific questions.

We only did this post-hoc analysis to add interpretability to the found effects with the semantic analysis; we show here the questions that showed a significant effect of ASD.

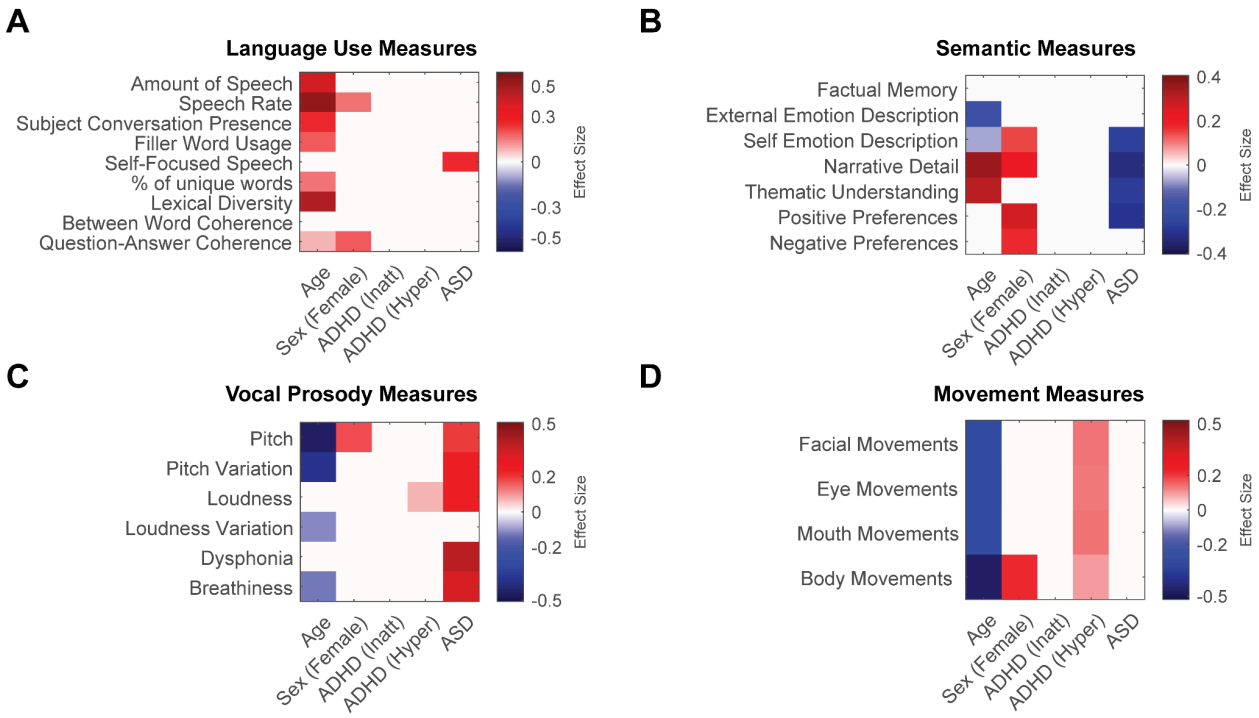

Figure S3. Results are independent of including IQ as a covariate.

Effect size matrices for primary behavioral outcomes after excluding IQ as a covariate. The models (A through D) are identical to the primary multivariate models presented in the main text, but without modelling IQ as a covariate. Effect sizes are estimated using multivariate regression, with $p<0.01$ after Bonferroni correction.

How do the features used here relate to established clinical instruments based on questionnaires? As one may expect, our measures of movement, language, and speech correlate strongly with these established metrics (Figure S4A, B). When using standard instruments, it appears that ADHD-Inattention is also associated with elevated motor activity (Fig. S4C). Thus, our movement measures appear to be more specifically associated with the ADHD-Hyperactive status (cf. Fig. 5C with S4C). On the other hand, when using standard instruments, it appears that ASD has a lesser effect on language use (Fig. S4D). Therefore, our features related to social communication appear to be more sensitive to the ASD presentation (cf. Fig. 3E with S4D).

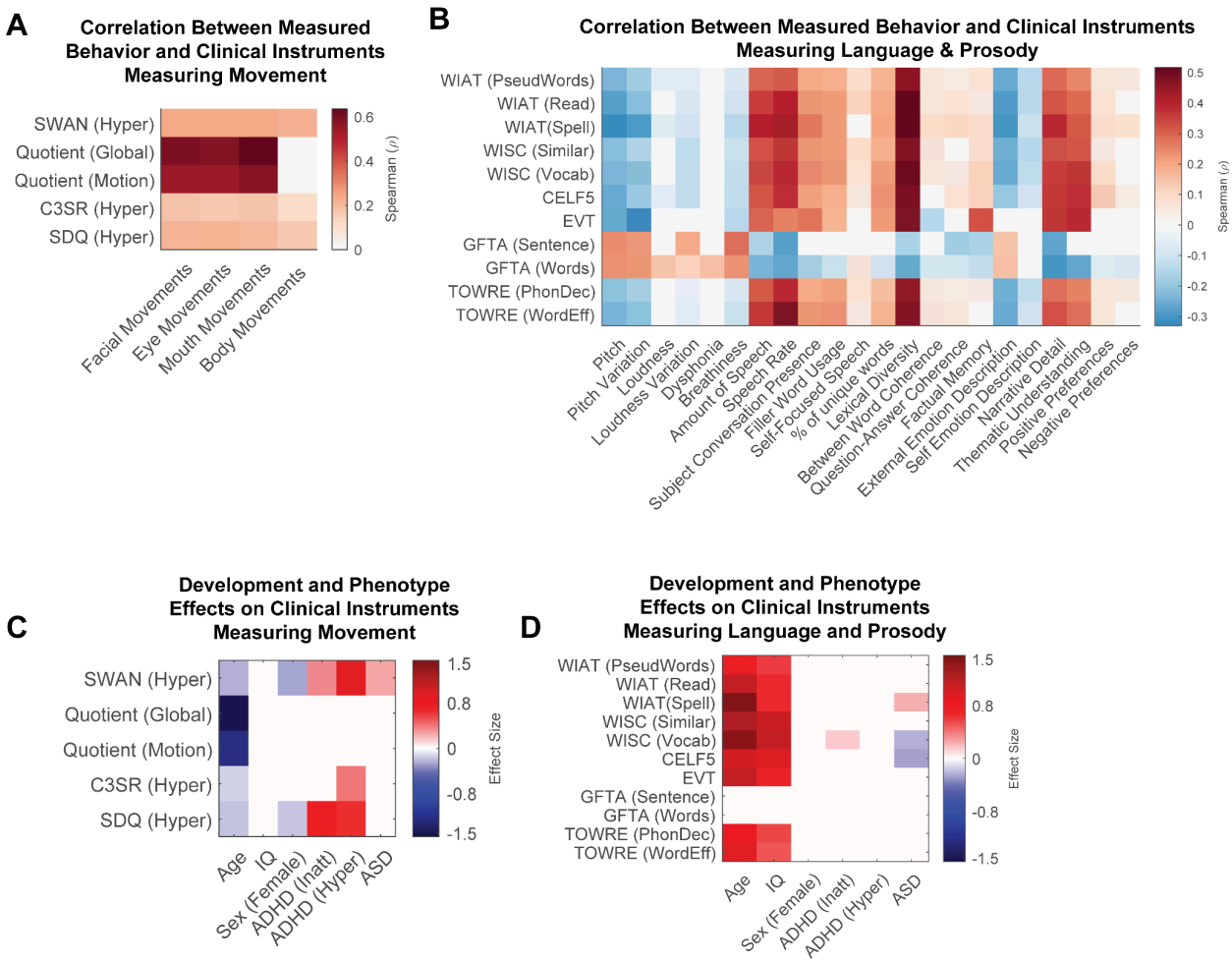

Figure S4. Comparing Clinical Instruments to measured behavior.

Spearman correlation between automatically measured behaviors and the clinical instrument scores measuring movement (A), and language use/prosody (B). C-D) Effects of development and diagnosis on such clinical instrument scores. Effect sizes in C, D measured using multivariate models, significance at $p<0.01$ after Bonferroni correction, white otherwise. For A and B, significant at a cutoff of $p<0.05$, uncorrected, white otherwise.

To select the most specific and robust trait measures for our analyses in Figures 1-2, we evaluated which of these scales best and most uniquely reflected the corresponding gold-standard, clinician-confirmed diagnosis. We ran separate multivariate regression models for each clinical instrument, entering the diagnoses as predictors. This approach allowed us to see the instruments most uniquely associated with each diagnosis. As shown here, the SWAN Inattention and Hyperactivity subscales and the ASSQ emerged as the strongest and most specific associations for their corresponding clinician-confirmed diagnoses. Therefore, we selected these scales for all subsequent analyses where continuous measures of ADHD and ASD traits were required:

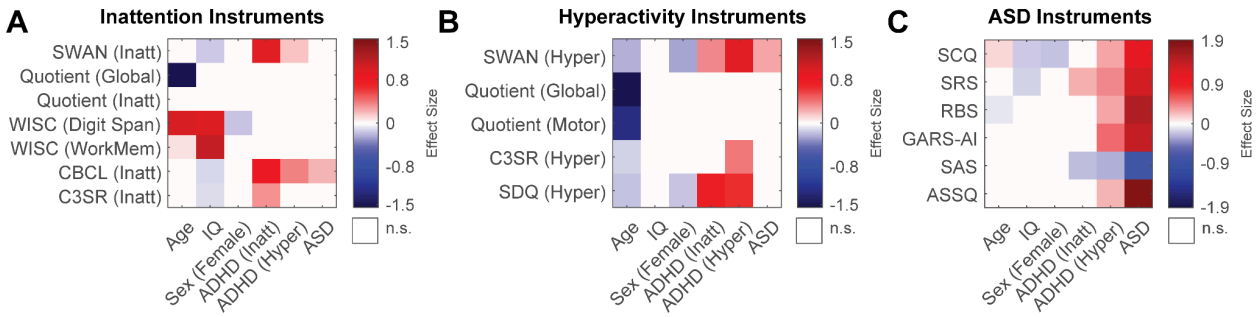

Figure S5. Unique association of clinical questionnaire scales with clinician-confirmed diagnoses.

The effect size matrices reveal which clinical instruments are more uniquely explained by specific positive diagnoses while accounting for development and demographic effects.

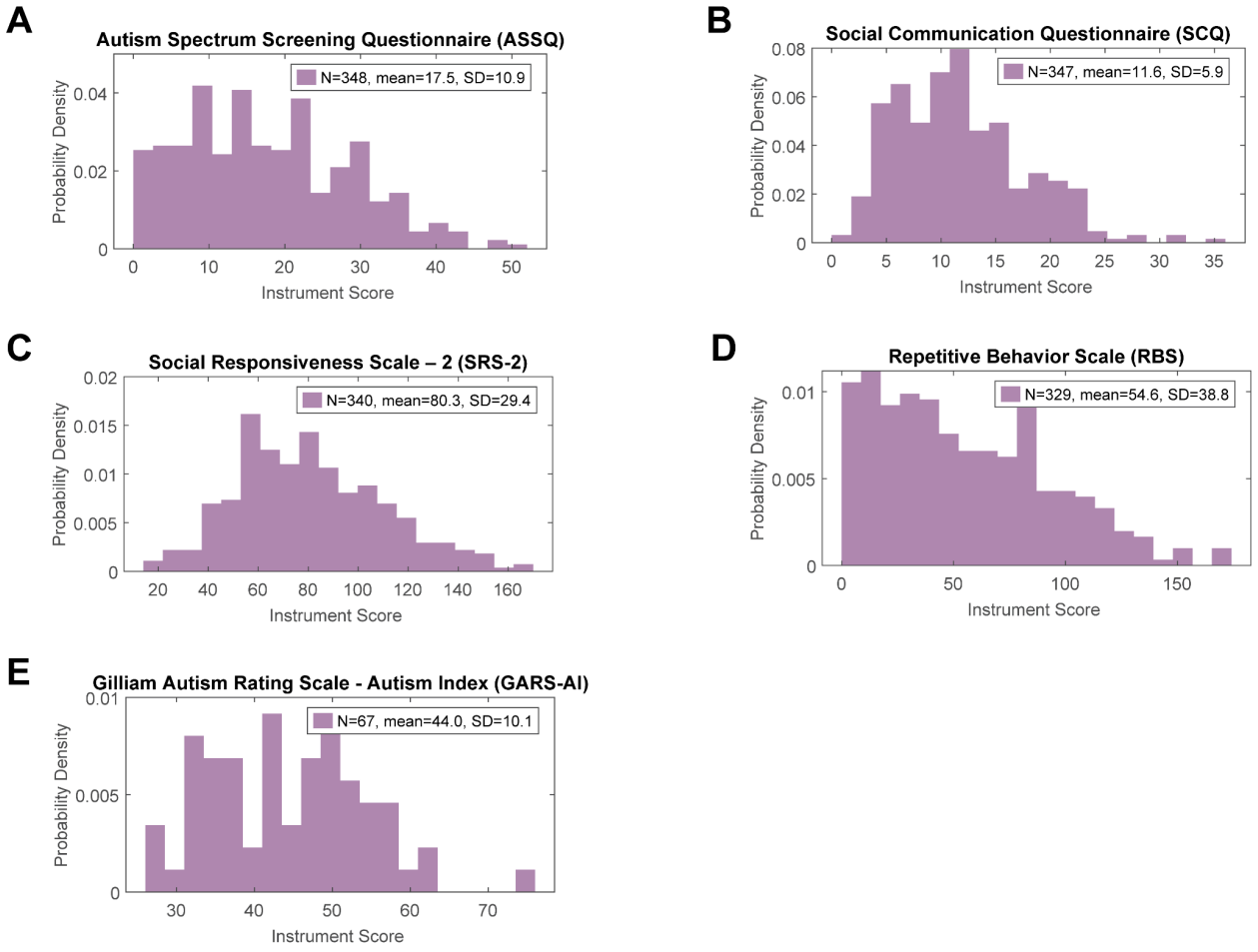

Figure S6. Distribution of Autism Scores in the ASD-positive population.

The Healthy Brain Network's deep phenotyping protocol provides multiple parent- and self-report questionnaires to measure traits associated with ADHD and ASD. These included the Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale (SWAN), the Autism Spectrum Screening

Questionnaire (ASSQ), the Social Communication Questionnaire (SCQ), the Social Responsiveness Scale-2 (SRS-2), the Child Behavior Checklist (CBCL), and the Conners 3 Self-Report (C3SR). The ASD population is mainly composed of mild-to-moderate cases.

[removed to meet bioRxiv policy regarding personal pictures]

Figure S7. Interview video examples, face landmarks, and validity.

The interviews were recorded in different locations and under different lighting conditions. However, all the videos showed the face and upper torso. **B)** 468 face landmarks tracked by Mediapipe. We characterized total movements for the whole face (blue), only the eyes (magenta), and the mouth (green). **C)** Most participants showed their face and torso for most part of the video (N=2108 and N=2226 participants were missing 10% or less of the data for face and pose, respectively); however, the amount of available data was less overall for the face landmarks than for the pose landmarks, as participants covered their face or looked away from the camera.

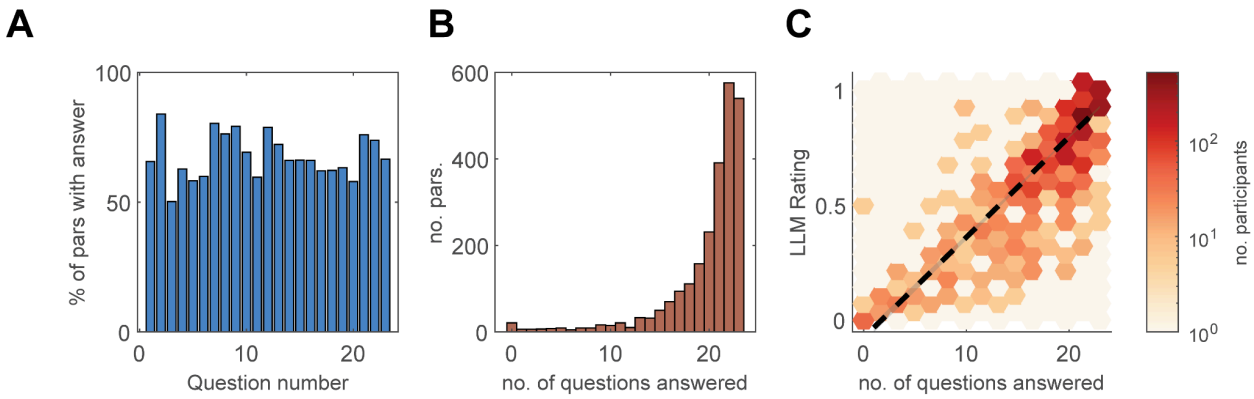

Figure S8. Validity of interview questions.

A) Not all questions were asked to every participant, or the answer was not available. In average, 67.7% of participants answered each question. **B)** Each participant answered 19.9 questions in average. **C)** The number of questions answered by each participant is correlated with the LLM's rating of the conversation completion ($\rho(2339) = 0.79$, $p<0.001$), see Supplementary Section S3 for rating prompt.

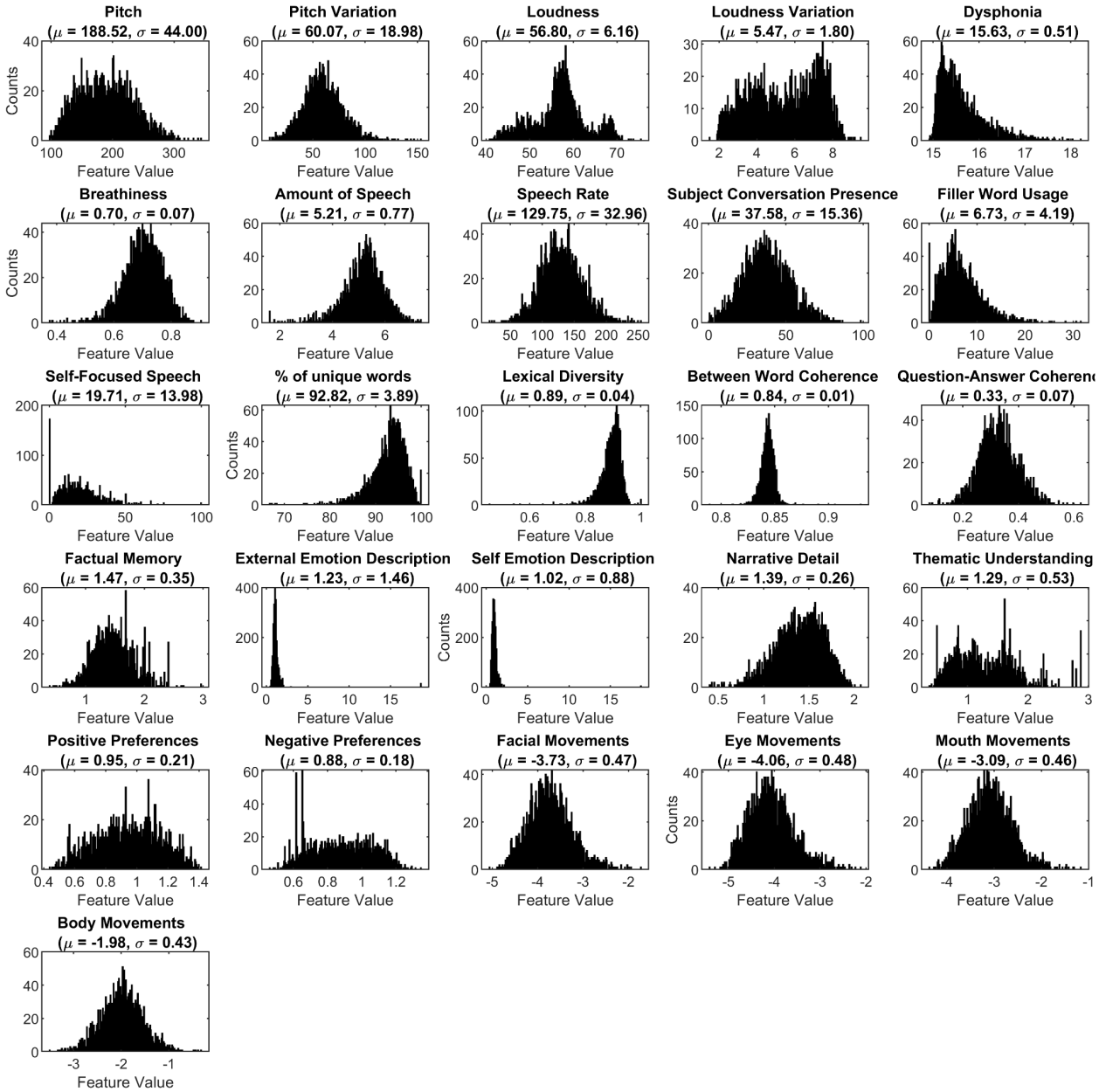

Figure S9. Distribution of Behavioral Variables.

A complete description of variables and their units is available in Table S3.

Given the high rates of anxiety in both ADHD and ASD, and the potential for a semi-structured interview with an unfamiliar clinician to elicit anxious behaviors, we modeled this as a key potential confound. Behaviors associated with anxiety, such as motor restlessness, social reticence, or altered vocal tone, could overlap with the objective behavioral markers under

investigation and potentially obscure the specific effects of ADHD and ASD. To address this and confirm the robustness of our primary findings, we repeated our primary multivariate regressions, adding a binary variable for any clinician-confirmed DSM-5 anxiety diagnosis as an additional covariate (N = 789). This allowed us to test whether the observed associations between our objective behavioral markers and ADHD/ASD diagnoses remained significant after statistically controlling for the influence of anxiety. We find that all the associations between ASD and/or ADHD with our behavioral measures are independent of whether we model anxiety or not:

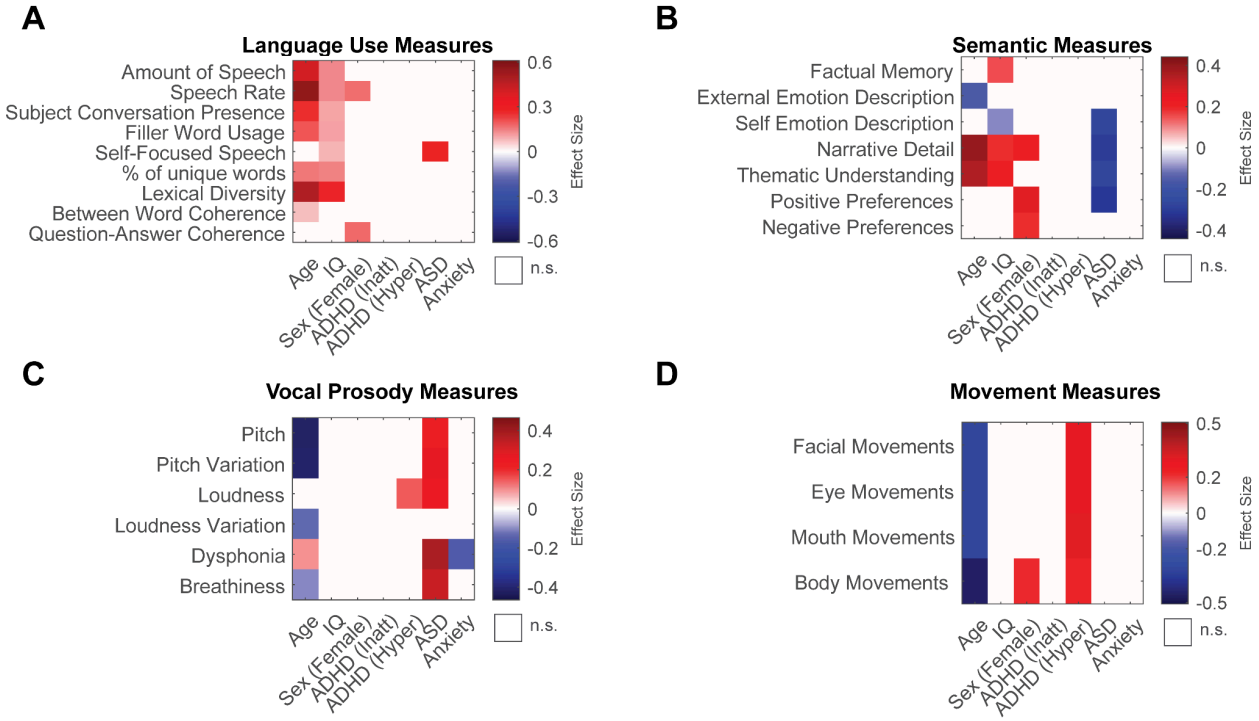

Figure S10. Results are independent of anxiety diagnoses.

Effect size matrices for primary behavioral outcomes after controlling for co-occurring anxiety. The models (A through D) are identical to the primary multivariate models presented in the main text, with the addition of a binary covariate for any clinician-confirmed anxiety diagnosis (N=789). Effect sizes are estimated using multivariate regression, with $p<0.01$ after Bonferroni correction.

Throughout the work presented in this manuscript, clinician-diagnosed ADHD-Combined presentations were modelled as positive in both ADHD-Hyperactive and ADHD-Inattentive status. This was done to maximize the number of participants in the modelled group, while maintaining the distinction of the two main inattentive and hyperactive diagnostic statuses. Previous studies on the HBN population have modelled this overlap by considering the ADHD-Combined separate from the ADHD-Inattentive presentations[97–99]. We repeated the analyses on the behavioral features by following this approach, and the results were mostly consistent our associations with the ADHD-Hyperactive presentation:

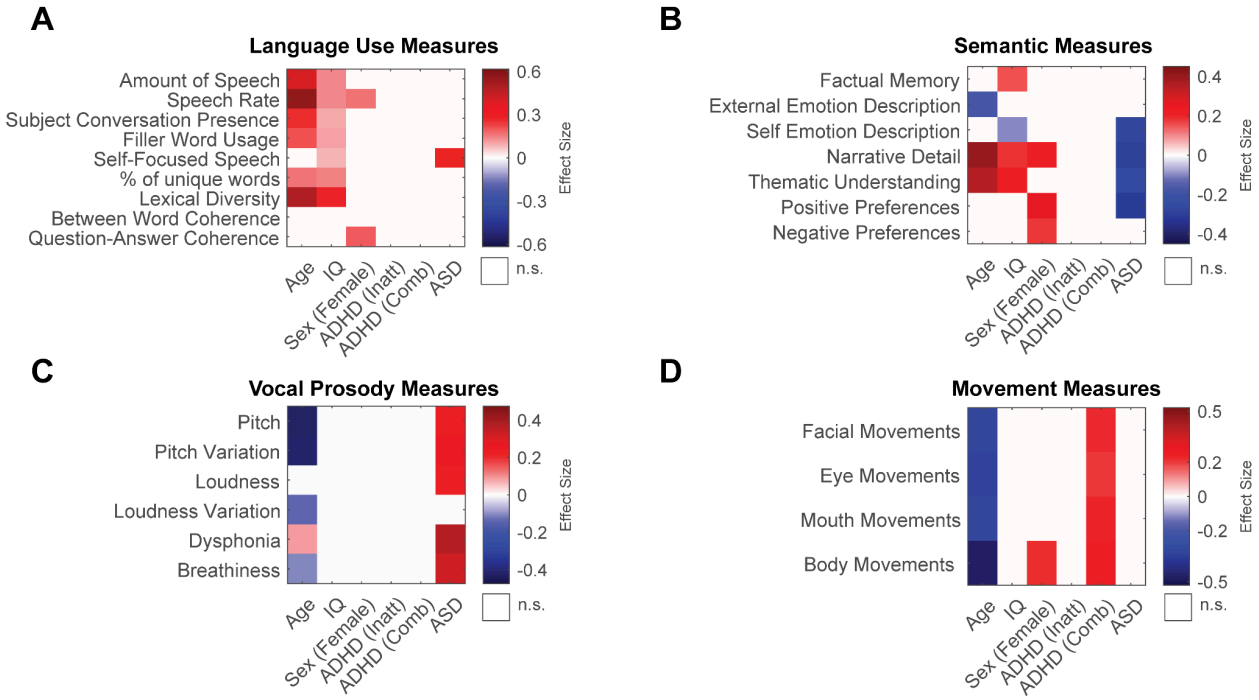

Figure S11. Results are independent of how the ADHD presentations are coded.

Effect size matrices for primary behavioral outcomes after modelling ADHD-Combined and ADHD-Inattentive presentations separately. The models (A through D) are identical to the primary multivariate models presented in the main text, but ADHD was modelled separately as the ADHD-Combined (N=625) and ADHD-Inattentive (N=653) presentations. Effect sizes are estimated using multivariate regression, with $p<0.01$ after Bonferroni correction.

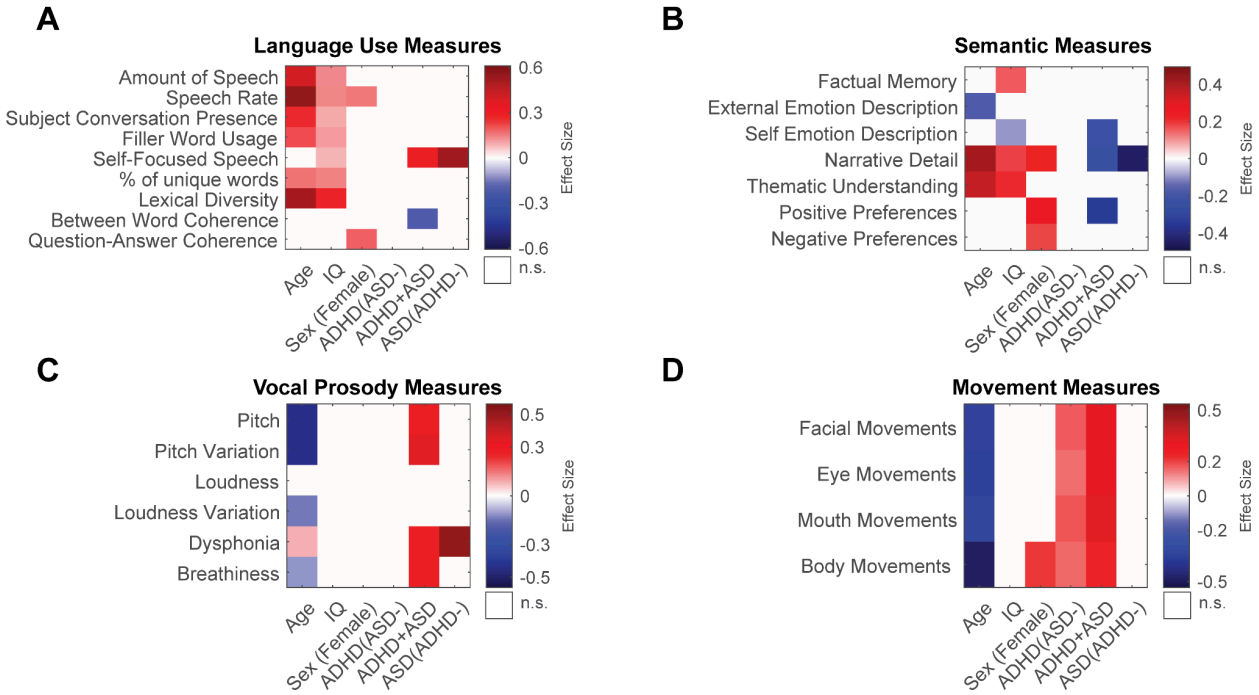

Figure S12. Modeling ASD(ADHD+) and ASD(ADHD-) separately does not paint a clear picture.

Effect size matrices for primary behavioral outcomes after treating co-occurring ADHD and ASD as a separate category. The models (A through D) are identical to the primary multivariate models presented in the main text, but without modelling IQ as a covariate. Effect sizes are estimated using multivariate regression, with p<0.01 after Bonferroni correction.

Supplementary Section S1. Multivariate Model Reports.

p-values reported in the tables below appear uncorrected for multiple comparisons unless otherwise specified. In the main manuscript, and in the reported corrected p-values below, all p-values are corrected for multiple comparisons using Bonferroni correction by multiplying these p-values by the number of behavioral features tested. Specifically, in Figures 2-5, Effect Size matrices show significant effect sizes in color with a cutoff of p<0.01 on the corrected p-value, and are left blank otherwise. See Methods for the explanation on analytical models.

<u>Supplementary Section S1.1. Diagnoses</u>

Detailed model reports for panel A in Figure 2 and black arrows in panel C in Figure 3.

--- Results for Outcome: ASD ---			
Predictor	Odds Ratio (OR)	95% CI	p-value
Age	1.033	[0.997, 1.07]	0.0746
Sex (Female)	0.426	[0.315, 0.575]	2.57E-08
ADHD (Inatt)	2.251	[1.69, 3]	3.55E-08
ADHD (Hyper)	1.597	[1.22, 2.08]	0.000548
--- Results for Outcome: ADHD (Inatt) ---			
Predictor	Odds Ratio (OR)	95% CI	p-value
Age	1.071	[1.04, 1.1]	1.03E-06
Sex (Female)	0.681	[0.56, 0.828]	0.000118
ADHD (Hyper)	10.008	[7.8, 12.8]	1.38E-73
ASD	2.256	[1.69, 3.01]	3.62E-08
--- Results for Outcome: ADHD (Hyper) ---			
Predictor	Odds Ratio (OR)	95% CI	p-value
Age	0.826	[0.797, 0.856]	7.84E-26
Sex (Female)	0.58	[0.46, 0.731]	3.90E-06
ADHD (Inatt)	10.46	[8.13, 13.5]	1.44E-74
ASD	1.601	[1.22, 2.09]	0.000593

Supplementary Section S1.2. Language Use Measures

Detailed model reports for panels D & E in Figure 2.

--- Results for Feature: Amount of Speech ---						
Model Fit: $R^2 = 0.154$, Adjusted $R^2 = 0.151$						
Overall Model: $F(6, 2104) = 63.662$, $p = 8.01e-73$						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	3.8	0.104	36.6	NaN	1.41E-227	1.27E-226
Age	0.0881	0.00483	18.3	0.398	2.98E-69	2.68E-68

IQ	0.00529	0.000845	6.25	0.136	4.87E-10	4.39E-09
Sex (Female)	0.0646	0.0302	2.14	0.0987	0.0327	0.294
ADHD (Inatt)	0.0435	0.0316	1.38	0.0603	0.169	1.52
ADHD (Hyper)	0.00959	0.0341	0.281	0.0131	0.779	7.01
ASD	0.00618	0.04	0.154	0.00941	0.877	7.9
--- Results for Feature: Speech Rate ---						
Model Fit: R^2 = 0.251, Adjusted R^2 = 0.249						
Overall Model: F(6, 2104) = 117.372, p = 4.34e-128						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	50.5	4.57	11	NaN	1.27E-27	1.15E-26
Age	5.49	0.212	25.8	0.563	6.74E-128	6.07E-127
IQ	0.231	0.0372	6.21	0.135	6.32E-10	5.69E-09
Sex (Female)	4.77	1.33	3.58	0.165	0.000349	0.00314
ADHD (Inatt)	-1.45	1.39	-1.04	-0.0457	0.297	2.67
ADHD (Hyper)	4.02	1.5	2.68	0.125	0.0075	0.0675
ASD	-3.28	1.76	-1.86	-0.113	0.063	0.567
--- Results for Feature: Subject Conversation Presence ---						
Model Fit: R^2 = 0.072, Adjusted R^2 = 0.069						
Overall Model: F(6, 2104) = 27.199, p = 2.19e-31						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	13.6	2.44	5.58	NaN	2.78E-08	2.50E-07
Age	1.33	0.113	11.7	0.256	8.63E-31	7.76E-30
IQ	0.0914	0.0199	4.6	0.1	4.51E-06	4.06E-05
Sex (Female)	1.11	0.711	1.56	0.072	0.119	1.07
ADHD (Inatt)	1.55	0.743	2.08	0.0911	0.0376	0.338
ADHD (Hyper)	0.0615	0.801	0.0768	0.00358	0.939	8.45
ASD	0.225	0.941	0.24	0.0146	0.811	7.3
--- Results for Feature: Filler Word Usage ---						

Model Fit: R^2 = 0.062, Adjusted R^2 = 0.060						
Overall Model: F(6, 2122) = 23.512, p = 4.96e-27						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	1.09	0.608	1.79	NaN	0.0741	0.667
Age	0.272	0.0282	9.65	0.21	1.33E-21	1.19E-20
IQ	0.0251	0.00496	5.06	0.11	4.60E-07	4.14E-06
Sex (Female)	0.0974	0.177	0.552	0.0253	0.581	5.23
ADHD (Inatt)	0.0397	0.185	0.215	0.00936	0.83	7.47
ADHD (Hyper)	0.111	0.199	0.559	0.0259	0.576	5.19
ASD	-0.388	0.233	-1.66	-0.1	0.0968	0.871
--- Results for Feature: Self-Focused Speech ---						
Model Fit: R^2 = 0.031, Adjusted R^2 = 0.029						
Overall Model: F(6, 2100) = 11.306, p = 1.83e-12						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	11.5	2.1	5.45	NaN	5.53E-08	4.98E-07
Age	-0.0402	0.0977	-0.411	-0.00897	0.681	6.13
IQ	0.0669	0.0171	3.9	0.0852	9.78E-05	0.000881
Sex (Female)	-1.23	0.613	-2	-0.0925	0.0456	0.41
ADHD (Inatt)	0.711	0.641	1.11	0.0486	0.267	2.41
ADHD (Hyper)	-0.313	0.691	-0.454	-0.0212	0.65	5.85
ASD	3.7	0.81	4.57	0.278	5.28E-06	4.75E-05
--- Results for Feature: % of unique words ---						
Model Fit: R^2 = 0.065, Adjusted R^2 = 0.062						
Overall Model: F(6, 2104) = 24.246, p = 6.8e-28						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	88.5	0.55	161	NaN	0	0
Age	0.188	0.0255	7.35	0.16	2.93E-13	2.63E-12
IQ	0.0298	0.00447	6.65	0.145	3.62E-11	3.26E-10
Sex (Female)	0.446	0.16	2.79	0.129	0.00538	0.0484

ADHD (Inatt)	-0.294	0.167	-1.76	-0.077	0.0788	0.709
ADHD (Hyper)	-0.243	0.18	-1.35	-0.063	0.177	1.6
ASD	0.0827	0.212	0.39	0.0238	0.696	6.27
--- Results for Feature: Lexical Diversity ---						
Model Fit: R^2 = 0.246, Adjusted R^2 = 0.244						
Overall Model: F(6, 2104) = 114.224, p = 5.05e-125						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	0.807	0.00429	188	NaN	0	0
Age	0.00446	0.000199	22.4	0.489	6.63E-100	5.97E-99
IQ	0.000437	3.49E-05	12.5	0.273	9.25E-35	8.32E-34
Sex (Female)	0.00295	0.00125	2.37	0.109	0.0181	0.163
ADHD (Inatt)	0.00242	0.00131	1.86	0.0813	0.0635	0.571
ADHD (Hyper)	-0.00282	0.00141	-2	-0.0935	0.0453	0.408
ASD	0.00149	0.00165	0.899	0.0548	0.369	3.32
--- Results for Feature: Between Word Coherence ---						
Model Fit: R^2 = 0.013, Adjusted R^2 = 0.010						
Overall Model: F(6, 2104) = 4.709, p = 9.03e-05						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	0.842	0.000926	910	NaN	0	0
Age	0.000141	4.30E-05	3.28	0.0714	0.00107	0.00966
IQ	1.99E-06	7.53E-06	0.264	0.00576	0.791	7.12
Sex (Female)	0.000214	0.000269	0.793	0.0366	0.428	3.85
ADHD (Inatt)	-0.000485	0.000282	-1.72	-0.0753	0.0858	0.772
ADHD (Hyper)	1.73E-05	0.000304	0.0568	0.00265	0.955	8.59
ASD	-0.00104	0.000357	-2.92	-0.178	0.00353	0.0318
--- Results for Feature: Question-Answer Coherence ---						
Model Fit: R^2 = 0.014, Adjusted R^2 = 0.011						
Overall Model: F(6, 2087) = 5.029, p = 3.96e-05						

Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	0.309	0.0113	27.3	NaN	1.06E-140	9.51E-140
Age	0.00151	0.000526	2.87	0.0628	0.00416	0.0375
IQ	-9.45E-06	9.20E-05	-0.103	-0.00225	0.918	8.26
Sex (Female)	0.0134	0.00329	4.08	0.189	4.73E-05	0.000426
ADHD (Inatt)	-0.00316	0.00344	-0.92	-0.0404	0.358	3.22
ADHD (Hyper)	0.00374	0.00371	1.01	0.0472	0.314	2.82
ASD	-0.0047	0.00435	-1.08	-0.0661	0.281	2.52

Supplementary Section S1.3. Semantic Measures

Detailed model reports for panel E in Figure 3.

--- Results for Feature: Factual Memory ---						
Model Fit: R^2 = 0.029, Adjusted R^2 = 0.026						
Overall Model: F(6, 2018) = 9.968, p = 7.2e-11						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	1.07	0.0588	18.3	NaN	4.73E-69	3.31E-68
Age	0.00601	0.00272	2.21	0.0491	2.74E-02	1.92E-01
IQ	0.0033	0.000479	6.89	0.153	7.46E-12	5.22E-11
Sex (Female)	-0.00183	0.0171	-0.107	-0.00503	0.915	6.41
ADHD (Inatt)	0.0372	0.0179	2.08	0.0929	0.0379	0.265
ADHD (Hyper)	-0.0283	0.0193	-1.47	-0.0701	0.142	0.995
ASD	-0.0366	0.0224	-1.63	-0.101	0.103	0.72
--- Results for Feature: External Emotion Description ---						
Model Fit: R^2 = 0.200, Adjusted R^2 = 0.198						
Overall Model: F(6, 2096) = 87.575, p = 3.43e-98						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	1.35	0.0551	24.5	NaN	8.47E-117	5.93E-116
Age	-0.0201	0.00255	-7.87	-0.172	5.58E-15	3.91E-14

IQ	-0.000759	0.000449	-1.69	-0.0369	9.09E-02	6.36E-01
Sex (Female)	-0.0129	0.016	-0.805	-0.0372	0.421	2.95
ADHD (Inatt)	0.0193	0.0167	1.15	0.0506	0.249	1.74
ADHD (Hyper)	0.00634	0.018	0.351	0.0164	0.725	5.08
ASD	-0.00759	0.0212	-0.358	-0.022	0.721	5.04
--- Results for Feature: Self Emotion Description ---						
Model Fit: R^2 = 0.080, Adjusted R^2 = 0.077						
Overall Model: F(6, 2080) = 30.197, p = 6.74e-35						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	1.17	0.0451	26	NaN	1.61E-129	1.13E-128
Age	-0.00568	0.00209	-2.72	-0.0597	6.51E-03	4.55E-02
IQ	-0.00163	0.000368	-4.44	-0.0973	9.51E-06	6.66E-05
Sex (Female)	0.0417	0.0131	3.18	0.148	0.00149	0.0104
ADHD (Inatt)	-0.00144	0.0137	-0.105	-0.00463	0.916	6.41
ADHD (Hyper)	0.0221	0.0148	1.5	0.0701	0.135	0.942
ASD	-0.0768	0.0174	-4.42	-0.272	1.06E-05	7.39E-05
--- Results for Feature: Narrative Detail ---						
Model Fit: R^2 = 0.181, Adjusted R^2 = 0.179						
Overall Model: F(6, 2084) = 76.996, p = 4.45e-87						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	0.78	0.0403	19.4	NaN	4.96E-77	3.47E-76
Age	0.0336	0.00187	18	0.394	3.00E-67	2.10E-66
IQ	0.00277	0.000327	8.46	0.185	4.76E-17	3.33E-16
Sex (Female)	0.0562	0.0117	4.82	0.223	1.52E-06	1.06E-05
ADHD (Inatt)	0.0228	0.0122	1.87	0.0823	0.0613	0.429
ADHD (Hyper)	-0.0272	0.0132	-2.06	-0.0969	0.0391	0.274
ASD	-0.0754	0.0154	-4.88	-0.299	1.15E-06	8.08E-06

--- Results for Feature: Thematic Understanding ---						
Model Fit: R^2 = 0.146, Adjusted R^2 = 0.143						
Overall Model: F(6, 1859) = 52.906, p = 2.14e-60						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	0.0751	0.0851	0.882	NaN	3.78E-01	2.64E+00
Age	0.0576	0.00387	14.9	0.345	1.73E-47	1.21E-46
IQ	0.0063	0.000689	9.14	0.212	1.55E-19	1.08E-18
Sex (Female)	-0.013	0.0241	-0.542	-0.0266	0.588	4.12
ADHD (Inatt)	0.031	0.0254	1.22	0.0569	0.222	1.56
ADHD (Hyper)	-0.0399	0.0274	-1.46	-0.0727	0.146	1.02
ASD	-0.131	0.032	-4.08	-0.266	4.65E-05	3.26E-04
--- Results for Feature: Positive Preferences ---						
Model Fit: R^2 = 0.043, Adjusted R^2 = 0.040						
Overall Model: F(6, 1867) = 14.105, p = 9.18e-16						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	0.844	0.0373	22.7	NaN	1.15E-100	8.04E-100
Age	0.00371	0.00172	2.16	0.05	3.07E-02	2.15E-01
IQ	0.000655	0.000304	2.16	0.0499	3.12E-02	2.19E-01
Sex (Female)	0.0587	0.0109	5.39	0.264	7.76E-08	5.43E-07
ADHD (Inatt)	-0.00798	0.0115	-0.697	-0.0324	0.486	3.4
ADHD (Hyper)	-0.0113	0.0123	-0.912	-0.0451	0.362	2.53
ASD	-0.0698	0.0144	-4.85	-0.313	1.31E-06	9.14E-06
--- Results for Feature: Negative Preferences ---						
Model Fit: R^2 = 0.014, Adjusted R^2 = 0.011						
Overall Model: F(6, 1798) = 4.358, p = 0.000224						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	0.851	0.0344	24.7	NaN	2.79E-116	1.96E-115
Age	0.00173	0.00158	1.1	0.0258	2.74E-01	1.91E+00

IQ	3.48E-05	2.79E-04	0.125	0.00294	9.01E-01	6.31E+00
Sex (Female)	0.0359	0.01	3.59	0.18	0.000345	0.00241
ADHD (Inatt)	-0.00515	0.0105	-0.491	-0.0233	0.624	4.37
ADHD (Hyper)	-0.0044	0.0113	-0.389	-0.0197	0.697	4.88
ASD	-0.0307	0.0131	-2.34	-0.153	0.0194	0.135

Supplementary Section S1.4. Vocal Prosody Measures

Detailed model reports for panel B in Figure 4.

--- Results for Feature: Pitch ---						
Model Fit: R^2 = 0.172, Adjusted R^2 = 0.170						
Overall Model: F(6, 2109) = 72.945, p = 7.27e-83						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	245	6.71	36.5	NaN	1.95E-226	1.17E-225
Age	-5.99	0.312	-19.2	-0.419	4.06E-76	2.44E-75
IQ	-0.016	0.0547	-0.292	-0.00636	0.77	4.62
Sex (Female)	4.33	1.95	2.22	0.102	0.0266	0.16
ADHD (Inatt)	0.198	2.04	0.0971	0.00425	0.923	5.54
ADHD (Hyper)	6.1	2.2	2.78	0.129	0.00554	0.0333
ASD	9.41	2.58	3.65	0.222	0.000271	0.00163
--- Results for Feature: Pitch Variation ---						
Model Fit: R^2 = 0.177, Adjusted R^2 = 0.174						
Overall Model: F(6, 2109) = 75.459, p = 1.52e-85						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	77.3	2.72	28.4	NaN	1.48E-150	8.87E-150
Age	-2.41	0.126	-19.1	-0.415	7.72E-75	4.63E-74
IQ	0.0318	0.0222	1.44	0.0313	0.151	0.906
Sex (Female)	1.47	0.791	1.86	0.0857	0.0633	0.38
ADHD (Inatt)	0.479	0.827	0.58	0.0254	0.562	3.37
ADHD (Hyper)	2.52	0.891	2.83	0.132	0.00472	0.0283
ASD	4.5	1.05	4.3	0.261	1.77E-05	0.000106

--- Results for Feature: Loudness ---						
Model Fit: R^2 = 0.021, Adjusted R^2 = 0.018						
Overall Model: F(6, 2109) = 7.474, p = 6.11e-08						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	58.2	0.992	58.6	NaN	0	0
Age	0.0745	0.0461	1.62	0.0352	0.106	0.636
IQ	-0.0169	0.00808	-2.09	-0.0456	0.0364	0.219
Sex (Female)	-0.684	0.288	-2.37	-0.109	0.0178	0.107
ADHD (Inatt)	-0.767	0.302	-2.54	-0.111	0.0111	0.0665
ADHD (Hyper)	1.05	0.325	3.23	0.15	0.00128	0.00765
ASD	1.45	0.381	3.8	0.231	0.000149	0.000892
--- Results for Feature: Loudness Variation ---						
Model Fit: R^2 = 0.029, Adjusted R^2 = 0.026						
Overall Model: F(6, 2109) = 10.384, p = 2.27e-11						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	6.85	0.308	22.2	NaN	1.54E-98	9.21E-98
Age	-0.0922	0.0143	-6.44	-0.14	1.45E-10	8.68E-10
IQ	-0.00428	0.00251	-1.7	-0.0371	0.0887	0.532
Sex (Female)	-0.274	0.0896	-3.06	-0.141	0.00226	0.0135
ADHD (Inatt)	0.0185	0.0937	0.197	0.00862	0.844	5.06
ADHD (Hyper)	-0.00554	0.101	-0.0549	-0.00255	0.956	5.74
ASD	0.26	0.118	2.2	0.134	0.0281	0.169
--- Results for Feature: Dysphonia ---						
Model Fit: R^2 = 0.052, Adjusted R^2 = 0.050						
Overall Model: F(6, 2109) = 19.360, p = 4.39e-22						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	15.5	0.0734	212	NaN	0	0

Age	0.0134	0.00341	3.94	0.0857	8.54E-05	0.000512
IQ	-0.00104	0.000598	-1.74	-0.0378	0.0824	0.494
Sex (Female)	-0.0243	0.0213	-1.14	-0.0524	0.255	1.53
ADHD (Inatt)	-0.0284	0.0223	-1.27	-0.0558	0.203	1.22
ADHD (Hyper)	0.0175	0.024	0.729	0.0339	0.466	2.8
ASD	0.172	0.0282	6.1	0.371	1.26E-09	7.56E-09
--- Results for Feature: Breathiness ---						
Model Fit: R^2 = 0.028, Adjusted R^2 = 0.025						
Overall Model: F(6, 2109) = 10.032, p = 5.94e-11						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	0.732	0.0107	68.6	NaN	0	0
Age	-0.00235	0.000495	-4.76	-0.104	2.11E-06	1.27E-05
IQ	-6.02E-05	8.68E-05	-0.694	-0.0151	0.488	2.93
Sex (Female)	0.00597	0.0031	1.93	0.0888	0.0541	0.325
ADHD (Inatt)	-0.00432	0.00324	-1.33	-0.0583	0.183	1.1
ADHD (Hyper)	0.00247	0.00349	0.707	0.0329	0.48	2.88
ASD	0.0219	0.00409	5.36	0.325	9.46E-08	5.68E-07

Supplementary Section S1.5. Movement Measures

Detailed model reports for panel B in Figure 5.

--- Results for Feature: Facial Movements ---						
Model Fit: R^2 = 0.136, Adjusted R^2 = 0.134						
Overall Model: F(6, 2141) = 56.376, p = 6.9e-65						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	-3.3	0.0697	-47.4	NaN	0	0
Age	-0.0474	0.00323	-14.7	-0.317	1.93E-46	7.71E-46
IQ	-0.000256	0.000568	-0.452	-0.00976	0.652	2.61
Sex (Female)	-0.026	0.0202	-1.29	-0.0587	0.199	0.795

ADHD (Inatt)	-0.0177	0.0212	-0.835	-0.0362	0.404	1.62
ADHD (Hyper)	0.147	0.0229	6.41	0.297	1.80E-10	7.21E-10
ASD	0.0739	0.0267	2.76	0.166	0.00575	0.023
--- Results for Feature: Eye Movements ---						
Model Fit: R^2 = 0.142, Adjusted R^2 = 0.140						
Overall Model: F(6, 2141) = 59.035, p = 7.78e-68						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	-3.62	0.0707	-51.2	NaN	0	0
Age	-0.0489	0.00328	-14.9	-0.322	7.28E-48	2.91E-47
IQ	-0.000178	0.000576	-0.309	-0.00669	0.757	3.03
Sex (Female)	-0.0551	0.0205	-2.69	-0.123	0.0073	0.0292
ADHD (Inatt)	-0.0224	0.0215	-1.04	-0.0452	0.297	1.19
ADHD (Hyper)	0.143	0.0232	6.14	0.284	1.01E-09	4.02E-09
ASD	0.0794	0.0271	2.93	0.176	0.00346	0.0138
--- Results for Feature: Mouth Movements ---						
Model Fit: R^2 = 0.130, Adjusted R^2 = 0.127						
Overall Model: F(6, 2141) = 53.259, p = 2.07e-61						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	-2.68	0.0678	-39.5	NaN	6.22E-257	2.49E-256
Age	-0.0449	0.00315	-14.3	-0.308	3.53E-44	1.41E-43
IQ	-0.000272	0.000553	-0.492	-0.0106	0.623	2.49
Sex (Female)	-0.00655	0.0197	-0.333	-0.0152	0.739	2.96

ADHD (Inatt)	-0.0135	0.0206	-0.654	-0.0284	0.513	2.05
ADHD (Hyper)	0.145	0.0223	6.49	0.301	1.03E-10	4.14E-10
ASD	0.0686	0.026	2.64	0.159	0.00845	0.0338
--- Results for Feature: Body Movements ---						
Model Fit: R^2 = 0.209, Adjusted R^2 = 0.207						
Overall Model: F(6, 2155) = 95.086, p = 3.18e-106						
Predictor	Coefficient (β)	Std. Error	t-stat	Effect Size	p-value	p-value (corrected)
(Intercept)	-1.33	0.0603	-22.1	NaN	6.80E-98	2.72E-97
Age	-0.06	0.0028	-21.4	-0.461	2.65E-92	1.06E-91
IQ	-0.0013	0.000492	-2.65	-0.057	0.00818	0.0327
Sex (Female)	0.0855	0.0175	4.88	0.222	1.14E-06	4.57E-06
ADHD (Inatt)	0.0139	0.0184	0.756	0.0327	0.45	1.8
ADHD (Hyper)	0.0967	0.0198	4.88	0.225	1.13E-06	4.51E-06
ASD	0.0259	0.0232	1.12	0.0671	0.264	1.06

Supplementary Section S2. Diarizing Prompt for LLM

```
""You are given a transcript of an interview between a clinician (interviewer) and a subject. The interviewer asks
questions from a predefined list, though slight variations may occur. Your task is to label the dialogue as either
"Interviewer:" or "Subject:", removing timestamps while preserving the dialogue structure.

### Instructions:
1. Identify lines spoken by the **interviewer** based on a predefined question list. These questions may have slight
variations but follow the same general intent.
2. Label all other responses as **"Subject:"**
3. Remove timestamps but keep the text structure intact.
4. Ensure proper formatting with a new line between each labeled turn.

### Example Input:

0.709 -> 2.83: I hope you enjoyed the last movie about the puppy.
2.95 -> 3.851: Have you seen it before?
4.451 -> 4.751: No.
5.231 -> 6.792: Can you tell me what happened in the movie?
6.852 -> 8.013: Try to tell the whole story.
8.093 -> 11.515: Remember that stories have a beginning, things that happen, and an ending.
12.895 -> 15.877: I didn't watch the whole thing though.
16.637 -> 18.478: Can you tell me about the part that you did watch?
19.139 -> 19.639: Okay.
```

```
### Example Output:

Interviewer: I hope you enjoyed the last movie about the puppy.
Interviewer: Have you seen it before?
Subject: No.
Interviewer: Can you tell me what happened in the movie?
Interviewer: Try to tell the whole story.
Interviewer: Remember that stories have a beginning, things that happen, and an ending.
Subject: I didn't watch the whole thing though.
Interviewer: Can you tell me about the part that you did watch?
Subject: Okay.

### If the transcript contains these specific sentences, you will label them as "clip" instead of interviewer or subject:

Whoa! Cool!
You gotta be kidding me
Get lost
Mom, I'll/we'll be outside

### The interviewer question list:

So I hope you enjoyed the last movie. Have you seen it before?
So can you tell me what happened in the movie? Try to tell the whole story. Remember that stories have a beginning,
things that happen, and an ending.
Do you remember anything else from the story?
So what are some of the things you liked about the movie?
What are some of the things you didn't like about the movie?
Who gave the boy a box?
What was in the box?
What was the boy doing before he got the box?
What was the puppy playing with?
How are the puppy and the boy the same?
So in the movie, who is missing a leg? The boy, the puppy, both the boy and the puppy, or no one?
So we're going to watch a short clip from the movie and then we'll talk about it.
How do you think the puppy was feeling?
How do you think the boy was feeling?
And how did you feel while you were watching that part?
How do you think the puppy was feeling?
How do you think the boy was feeling?
And how did you feel while you were watching that part?
How do you think the puppy was feeling?
How do you think the boy was feeling?
And how did you feel while you were watching that part?
How do you think the puppy was feeling?
How do you think the boy was feeling?
And how did you feel while you were watching this part?
Great, thank you

### In some cases, the interviewer may ask additional questions or make comments. You should label these as
"Interviewer:" as well using their specific context.

### The transcript:
<<transcript>>
```

Supplementary Section S3. Question-Answer Extraction Prompt for LLM

```
"""Task: Extract Answers from Interview Transcript

You are tasked with extracting the participant's answers to a specific set of predefined questions from an interview
transcript. The questions relate to the short film "The Present." The interviewer *may* have asked these questions
verbatim, asked them in a slightly different way, or not asked them at all. The subject may have answered directly,
indirectly, partially, over multiple turns, or not at all.
In the transcript the interviewer asks these eleven questions in order:
1.  "I hope you enjoyed the last movie. Have you seen it before?"
2.  "Can you tell me what happened in the movie? Try to tell the whole story. Remember that stories have a beginning,
things that happen, and an ending."
3.  "Do you remember anything else from the story?"
4.  "What are some of the things you liked about the movie?"
5.  "What are some of the things you didn't like about the movie?"
6.  "Who gave the boy a box?"
7.  "What was in the box?"
8.  "What was the boy doing before he got the box?"
9.  "What was the puppy playing with?"
10. "How are the puppy and the boy the same?"
11. "In the movie, who is missing a leg? The boy, the puppy, both the boy and the puppy, or no one?"

After these questions, the interviewer then shows the subject four short clips from the movie and asks the subject to
describe how the characters are feeling in each clip. The interviewer may say something like:
- "Let's watch a short clip from the movie, and then we'll talk about it."
- "Now let's watch another clip"
- "One more clip"
- "One last clip"

After watching each clip, the interviewer asks the subject how they think the characters are feeling and how they
themselves feel while watching the clip. The interviewer may ask these questions in various ways, such as:
- "How do you think the puppy was feeling?"
- "How do you think the boy was feeling?"
- "And how did you feel while you were watching that part?"


Input:

*   A transcript of a conversation between an interviewer and a subject discussing the short film "The Present."

Output:

Your output will consist of two parts:

**Part 1: Question-Answer Extraction**

A series of lines, one for each predefined question. Each line will follow this format:

`"Question" -> "Answer"`

Where:

*   `"Question"` is one of the predefined questions listed below (use the *exact* wording provided here, even if the
interviewer phrased it differently).
*   `"Answer"` is the subject's answer to the question.
    *   If the question was answered (directly or indirectly), provide the subject's complete answer, concatenating all
relevant parts of their response, even if it spans multiple turns or requires piecing together information from different
parts of the transcript.
```

```
    *   If the question was *not* asked, or if the subject provided *no* discernible answer (even after considering the
surrounding context), leave the "Answer" field *blank*. Do *not* write "N/A," "No answer," or any other placeholder. Just
leave it blank.
    * If parts of different answers are mixed, extract only the answer to the intended question.

Predefined Questions:

1.  "I hope you enjoyed the last movie. Have you seen it before?"
2.  "Can you tell me what happened in the movie? Try to tell the whole story. Remember that stories have a beginning,
things that happen, and an ending."
3.  "Do you remember anything else from the story?"
4.  "What are some of the things you liked about the movie?"
5.  "What are some of the things you didn't like about the movie?"
6.  "Who gave the boy a box?"
7.  "What was in the box?"
8.  "What was the boy doing before he got the box?"
9.  "What was the puppy playing with?"
10. "How are the puppy and the boy the same?"
11. "In the movie, who is missing a leg? The boy, the puppy, both the boy and the puppy, or no one?"
12. "How do you think the puppy was feeling? (after watching first clip)"
13. "How do you think the boy was feeling? (after watching first clip)"
14. "And how did you feel while you were watching that part? (after watching first clip)"
15. "How do you think the puppy was feeling? (after watching second clip)"
16. "How do you think the boy was feeling? (after watching second clip)"
17. "And how did you feel while you were watching that part? (after watching second clip)"
18. "How do you think the puppy was feeling? (after watching third clip)"
19. "How do you think the boy was feeling? (after watching third clip)"
20. "And how did you feel while you were watching that part? (after watching third clip)"
21. "How do you think the puppy was feeling? (after watching fourth clip)"
22. "How do you think the boy was feeling? (after watching fourth clip)"
23. "And how did you feel while you were watching that part? (after watching fourth clip)"

**Part 2: Conversation Completion Rating**

A single line in the following format:

`"Conversation Completion Rating" -> "Rating"`

Where:

*   `"Rating"` is a numerical score between 0 and 1 (inclusive) representing the overall completion of the conversation,
*specifically with respect to the predefined questions*.
    *   **1:** Represents a perfect completion, where all of the predefined questions were asked (or closely
paraphrased), and the subject provided clear and complete answers to each.
    *   **0:** Represents a very poor completion, where many questions were not asked, or the subject consistently failed
to answer the questions that were asked.
    *   **Intermediate values:** Represent varying degrees of completion quality.  A score of 0.5, for example, would
suggest that roughly half of the questions were asked and answered reasonably well. Use your judgment to assign a score
that reflects the overall completeness and clarity of the question-answer exchange related to the *predefined questions*.
Do *not* assess the overall quality of the conversation beyond its adherence to addressing the core content of these
questions.

Instructions:

1.  **Read the Entire Transcript:** Carefully read the entire transcript to understand the completion of the conversation
and the context of each statement.
2.  **Identify Questions and Answers:** For each predefined question:
```

```
    *   **Search for the Question:** Look for the exact question, or a close paraphrase, in the interviewer's speech.
Note that the interviewer might rephrase, prompt, or interrupt the subject.
    *   **Locate the Answer (If Present):** If the question (or a close variant) was asked, carefully examine the
subject's subsequent responses. The answer might be:
        *   **Direct and Immediate:** Right after the question.
        *   **Indirect:** Implied by their statements, requiring inference.
        *   **Partial:** Only part of the answer is given directly.
        *   **Scattered:** Pieces of the answer are given across multiple turns, possibly with interviewer prompts or
interruptions.
        *   **Non-existent:** The subject might not answer at all.
    * **Use Context:** Consider surrounding conversational turns from both interviewer and subject to determine if a
particular utterance constitutes an answer, even an indirect or partial one.
    * **Combine Response Parts:** If the answer is spread over multiple turns, concatenate *all relevant parts* of the
subject's response into a single, coherent answer. Remove any interviewer interjections or prompts from within the
concatenated answer. Only include the *subject's* words in the "Answer" field.
4.  * **For the questions after the clips:** Because there are multiple clips, you will need to extract the answers to
these questions from the conversation after each clip is shown. The interviewer will ask the subject how they think the
characters are feeling and how they themselves feel while watching the clip. However, the order of the clips is always
the same. Sometimes part of the dialogue is present on the transcript, so in those cases, use that as clue to know to
what clip the question is referring to. If you can't clearly know what question/answer corresponds to what clip, restrain
from returning an answer for that specific question, leave the "Answer" field blank.:
    * **Clip 1:**: the kid opens the present and sees the puppy, the transcript might say "Whoa, cool."
    * **Clip 2:**: the kid throws the puppy away, the transcript might say "You've got to be kidding me."
    * **Clip 3:**: the kid kicks the puppy, the transcript might say "Get lost!"
    * **Clip 4:**: the kid plays with the puppy, the transcript might say "Mom, I'll / We'll be outside."
3.  **Output (Part 1):** Generate the question-answer extraction in the specified `"Question" -> "Answer"` format. Leave
the "Answer" field blank if the question was not asked or if no answer can be found.
4. **Assess Conversation Completion (Part 2):** After extracting the question-answer pairs, evaluate the *overall
completion* of the conversation *with respect to the predefined questions*. Consider:
    *  How many of the predefined questions (or close paraphrases) were asked by the interviewer?
    *  Did the conversation stay focused on the topics covered by the predefined questions, or did it frequently deviate?
5.  **Output (Part 2):** Provide the "Conversation Completion Rating" on a separate line, using a number between 0 and 1.

Example 1:

    (example 1 transcript):
        Interviewer: So I hope you enjoyed the last movie.
        Interviewer: Have you seen it before?
        Subject: I have at the mock MRI.
        Interviewer: Oh, cool.
        Subject: I liked it, though.
        Interviewer: Great.
        Interviewer: So can you tell me what happened in the movie?
        Interviewer: Try to tell the whole story.
        Interviewer: Remember that stories have a beginning, things that happen, and an ending.
        Subject: So there's a boy who was disabled, and he didn't want to go outside.
        Subject: He wasn't in the mood for anything except for his video games.
        Subject: and then like his mom was trying to like help him go like get some fresh air and then but like
everything like his mom tried to do didn't work so then he then she bought um him a puppy so then he started like he
didn't want like to see the puppy like anymore because then he realized he realized that he was also like disabled
        Subject: But then he started having a liking to him because he felt like the puppy and him were like the same.
        Subject: So he started playing with the puppy more.
        Subject: So then they went outside and they started playing with the ball.
        Interviewer: Great.
        Interviewer: Do you remember anything else from the story?
        Subject: I think that's much it.
        Interviewer: Great.
        Interviewer: So what are some of the things you liked about the movie?
```

```
        Subject: I liked how he had a good relationship with his doll.
        Interviewer: What are some of the things you didn't like about the movie?
        Subject: I liked everything about the movie, but
        Subject: I was hoping it would be longer.
        Interviewer: Who gave the boy a box?
        Subject: His mother.
        Interviewer: What was in the box?
        Subject: The puppy.
        Interviewer: What was the boy doing before he got the box?
        Subject: He wasn't really paying attention to anything.
        Subject: He was glued to his video games.
        Interviewer: What was the puppy playing with?
        Subject: He was playing with the red rubber ball.
        Interviewer: How are the puppy and the boy the same?
        Subject: Both of them each have a disabled part of their body.
        Interviewer: So in the movie, who is missing a leg?
        Interviewer: The boy, the puppy, both the boy and the puppy, or no one?
        Subject: The boy.
        Interviewer: Great.
        Interviewer: So we're going to watch a short clip from the movie and then we'll talk about it.
        Subject: Okay.
        clip: Whoa, cool.
        Interviewer: How do you think the puppy was feeling?
        Subject: Happy.
        Interviewer: How do you think the boy was feeling?
        Subject: He's feeling happy too.
        Interviewer: And how did you feel while you were watching that pot?
        Subject: I guess happy.
        Interviewer: Right.
        Interviewer: How do you think the puppy was feeling?
        Subject: Sad.
        Interviewer: How do you think the boy was feeling?
        Subject: He was feeling angry.
        Interviewer: And how did you feel while you were watching that mom?
        Subject: I didn't like it.
        Interviewer: How do you think the puppy was feeling?
        Subject: He didn't mind.
        Interviewer: How do you think the boy was feeling?
        Subject: He was still angry.
        Interviewer: And how did you feel while you were watching that pup?
        Subject: I feel bad for the puppy.
        Interviewer: Great.
        clip: Mom, I'll be outside.
        Interviewer: How do you think the puppy was feeling?
        Subject: Um, happy.
        Interviewer: How do you think the boy was feeling?
        Subject: He was feeling happy too.
        Interviewer: And how did you feel while you were watching this pot?
        Subject: I was happy.
        Interviewer: Great, thank you.

    (example 1 output):
        "I hope you enjoyed the last movie. Have you seen it before?" -> "I have at the mug MRI."
        "Can you tell me what happened in the movie? Try to tell the whole story. Remember that stories have a beginning,
things that happen, and an ending." -> "So there='s a boy who was disabled, and he didn't want to go outside. He wasn't
in the mood for anything except for his video games. And then like his mom was trying to like help him go like get some
fresh air and then but like everything like his mom tried to do didn't work so then he then she bought um him a puppy so
then he started like he didn't want like to see the puppy like anymore because then he realized he realized that he was
```

also like disabled. But then he started having a liking to him because he felt like the puppy and him were like the same. So he started playing with the puppy more. So then they went outside and they started playing with the ball."

"Do you remember anything else from the story?" -> "I think that's much it."

"What are some of the things you liked about the movie?" -> "I liked how he had a good relationship with his doll."

"What are some of the things you didn't like about the movie?" -> "I liked everything about the movie, but I was hoping it would be longer."

"Who gave the boy a box" -> "His mother."

"What was in the box?" -> "The puppy."

"What was the boy doing before he got the box?" -> "He wasn't really paying attention to anything. He was glued to his video games."

"What was the puppy playing with?" -> "He was playing with the red rubber ball."

"How are the puppy and the boy the same?" -> "Both of them each have a disabled part of their body."

"In the movie, who is missing a leg? The boy, the puppy, both the boy and the puppy, or no one?" -> "The boy."

"How do you think the puppy was feeling? (after watching first clip)" -> "Happy."

"How do you think the boy was feeling? (after watching first clip)" -> "He's feeling happy too."

"And how did you feel while you were watching that part? (after watching first clip)" -> "I guess happy."

"How do you think the puppy was feeling? (after watching second clip)" -> "Sad."

"How do you think the boy was feeling? (after watching second clip)" -> "He was feeling angry."

"And how did you feel while you were watching that part? (after watching second clip)" -> "I didn't like it."

"How do you think the puppy was feeling? (after watching third clip)" -> "He didn't mind."

"How do you think the boy was feeling? (after watching third clip)" -> "He was still angry."

"And how did you feel while you were watching that part? (after watching third clip)" -> "I feel bad for the puppy"

"How do you think the puppy was feeling? (after watching fourth clip)" -> "Um, happy."

"How do you think the boy was feeling? (after watching fourth clip)" -> "He was feeling happy too."

"And how did you feel while you were watching that part? (after watching fourth clip)" -> "I was happy."

"Rating" -> "1"

Example 2:

(example 2 transcript):

Interviewer: Okay, so I hope you enjoyed the last movie.

Interviewer: Have you seen it before?

Subject: No

Interviewer: Um, so I'm going to need you to talk a little bit.

Interviewer: Is that okay?

Subject: Okay.

Interviewer: Can you tell me what happened in the movie?

Interviewer: Try to tell the whole story.

Interviewer: Remember that stories have a beginning, things that happen, and an ending.

Interviewer: Great.

Interviewer: Do you remember anything else from the story?

Interviewer: Okay.

Interviewer: What are some of the things you liked about the movie?

Interviewer: And what are some of the things you didn't like about the movie?

Interviewer: So who gave the boy a box?

Interviewer: What was in the box?

Interviewer: What was the boy doing before he got the box?

Interviewer: What was the puppy playing with?

Interviewer: How are the puppy and the boy the same?

Interviewer: Perfect.

Interviewer: So in the movie, who is missing a leg?

Interviewer: The boy, the puppy, both the boy and the puppy, or no one?

Subject: Both the boy and the puppy.

Interviewer: Great.

Interviewer: So we're going to watch a couple of short clips from the movie and talk about them, OK?

Subject: OK.

```
        clip: Whoa.
        clip: Cool.
        Interviewer: How do you think the puppy was feeling?
        Subject: Happy.
        Interviewer: How do you think the boy was feeling?
        Subject: Happy.
        Interviewer: Happy?
        Interviewer: And how did you feel while you were watching this part?
        Subject: Good.
        Interviewer: Good?
        Interviewer: Great.
        clip: You've got to be kidding me.
        Interviewer: How do you think the puppy was feeling?
        Interviewer: And how do you think the boy was feeling?
        Interviewer: And how did you feel while you were watching this part?
        clip: Get lost!
        Interviewer: How do you think the puppy was feeling?
        Interviewer: And how do you think the boy was feeling?
        Interviewer: And how did you feel while you were watching this pot?
        Interviewer: Okay, one more.
        clip: Mom!
        clip: We'll be outside!
        Interviewer: How do you think the puppy was feeling?
        Subject: Happy.
        Interviewer: How do you think the boy was feeling?
        Subject: Happy.
        Interviewer: And how did you feel while you were watching his paw?
        Subject: Happy.
        Interviewer: Great, thank you.


    (example 2 output):
        "I hope you enjoyed the last movie. Have you seen it before?" -> "No."
        "Can you tell me what happened in the movie? Try to tell the whole story. Remember that stories have a beginning,
things that happen, and an ending." -> ""
        "Do you remember anything else from the story?" -> ""
        "What are some of the things you liked about the movie?" -> ""
        "What are some of the things you didn't like about the movie?" -> ""
        "Who gave the boy a box" -> ""
        "What was in the box?" -> ""
        "What was the boy doing before he got the box?" -> ""
        "What was the puppy playing with?" -> ""
        "How are the puppy and the boy the same?" -> ""
        "In the movie, who is missing a leg? The boy, the puppy, both the boy and the puppy, or no one?" ->  "Both the
boy and the puppy."
        "How do you think the puppy was feeling? (after watching first clip)" -> "Happy."
        "How do you think the boy was feeling? (after watching first clip)" -> "Happy."
        "And how did you feel while you were watching that part? (after watching first clip)" -> "Good."
        "How do you think the puppy was feeling? (after watching second clip)" -> ""
        "How do you think the boy was feeling? (after watching second clip)" -> ""
        "And how did you feel while you were watching that part? (after watching second clip)" -> ""
        "How do you think the puppy was feeling? (after watching third clip)" -> ""
        "How do you think the boy was feeling? (after watching third clip)" -> ""
        "And how did you feel while you were watching that part? (after watching third clip)" -> ""
        "How do you think the puppy was feeling? (after watching fourth clip)" -> "Happy."
        "How do you think the boy was feeling? (after watching fourth clip)" -> "Happy."
        "And how did you feel while you were watching that part? (after watching fourth clip)" -> "Happy."
        "Rating" -> "0.1"
```

```
    (example 3 transcript):
        Interviewer: Have you seen that last movie before, the puppy one?
        Subject: I don't think I have.
        Interviewer: Can you try and tell me what happens in the movie?
        Interviewer: The full beginning, middle, and end.
        Subject: Um, the beginning, the beginning of the movie is, um, so it's the middle, mom gives him the puppy, but
he doesn't like it on first because it has three legs, and then towards the end,
        Subject: The last time I played with them.
        Subject: I don't know about it.
        Subject: All right, we both have the same disability.
        Subject: Okay.
        Subject: Didn't go home.
        Subject: How deep?
        Interviewer: What was the boy doing before he got the box?
        Interviewer: What was the puppy playing with?
        Interviewer: How are the puppy and the boy the same?
        Interviewer: In the movie, who's missing a leg?
        Interviewer: The boy, the puppy, both the boy and the puppy or no one?
        Interviewer: Okay, now we're going to watch a short clip and then we'll talk about it.
        Interviewer: How's the puppy feeling here?
        Interviewer: How's the boy feeling?
        Interviewer: And how do you feel when you watch his part?
        Interviewer: How's the puppy feeling here?
        Interviewer: How's the boy feeling?
        Interviewer: And how do you feel when you watch this part?
        Interviewer: How's the puppy feeling here?
        Interviewer: How's the boy feeling?
        Interviewer: How do you feel when you watch his fart?
        Interviewer: How's the puppy feeling here?
        Interviewer: How's the boy feeling?
        Interviewer: And how do you feeling off that part?
        Subject: Good.
        Interviewer: Okay, we're all done.

    (example 3 output):
        "I hope you enjoyed the last movie. Have you seen it before?" -> "I don't think I have."
        "Can you tell me what happened in the movie? Try to tell the whole story. Remember that stories have a beginning,
things that happen, and an ending." -> "Um, the beginning, the beginning of the movie is, um, so it's the middle, mom
gives him the puppy, but he doesn't like it on first because it has three legs, and then towards the end, The last time I
played with them. I don't know about it. All right, we both have the same disability. Okay. Didn't go home. How deep?"
        "Do you remember anything else from the story?" -> ""
        "What are some of the things you liked about the movie?" -> ""
        "What are some of the things you didn't like about the movie?" -> ""
        "Who gave the boy a box" -> ""
        "What was in the box?" -> ""
        "What was the boy doing before he got the box?" -> ""
        "What was the puppy playing with?" -> ""
        "How are the puppy and the boy the same?" -> ""
        "In the movie, who is missing a leg? The boy, the puppy, both the boy and the puppy, or no one?" ->  ""
        "How do you think the puppy was feeling? (after watching first clip)" -> ""
        "How do you think the boy was feeling? (after watching first clip)" -> ""
        "And how did you feel while you were watching that part? (after watching first clip)" -> ""
        "How do you think the puppy was feeling? (after watching second clip)" -> ""
        "How do you think the boy was feeling? (after watching second clip)" -> ""
        "And how did you feel while you were watching that part? (after watching second clip)" -> ""
        "How do you think the puppy was feeling? (after watching third clip)" -> ""
        "How do you think the boy was feeling? (after watching third clip)" -> ""
```

```
    "And how did you feel while you were watching that part? (after watching third clip)" -> ""
    "How do you think the puppy was feeling? (after watching fourth clip)" -> ""
    "How do you think the boy was feeling? (after watching fourth clip)" -> ""
    "And how did you feel while you were watching that part? (after watching fourth clip)" -> "Good."
    "Rating" -> "0"

(example 4 transcript):
    Interviewer: I hope you enjoyed the last movie.
    Interviewer: Have you seen it before?
    Subject: Yeah.
    Subject: The puppy cartoon?
    Subject: Yeah.
    Interviewer: Can you tell me what happened in the movie and try to tell the whole story?
    Interviewer: Remember that stories have a beginning, things that happen, and an ending.
    Subject: So first there was a boy that was playing a game.
    Subject: So when his mother came in the door was a box.
    Subject: She put it down and then a puppy was in the box.
    Subject: But the boy didn't want to play.
    Subject: Instead he wanted to play on his TV.
    Subject: So first he just kicked the puppy, so then he just came back and he did it again.
    Subject: So then he just went crazy and then got a ball.
    Subject: And then he wanted to play, but then the boy just kicked the ball in the box and the puppy ran with it.
    Subject: And then
    Subject: he dropped the ball back.
    Subject: And then the boy stopped when he was playing, and then he went to lose the puppy.
    Interviewer: Do you remember anything else from the story?
    Subject: That's all I can remember.
    Interviewer: What are some of the things you liked about the movie?
    Subject: When you like
    Subject: And then all of the puppies ran with it and then he got trapped in the box.
    Subject: When he kicked the puppy.
    Subject: His mom.
    Subject: A puppy.
    Subject: Playing on his TV.
    Subject: A ball.
    Interviewer: How are the puppy and the boy the same?
    Subject: They're both playing with them.
    Interviewer: In the movie, who is missing a leg?
    Interviewer: The boy, the puppy, both the boy and the puppy, or no one?
    Subject: The boy and the puppy.
    Interviewer: Let's watch a short clip from the movie, and then we'll talk about it.
    Interviewer: Give me one second, bud.
    Subject: Can we do an interview on it?
    Interviewer: This is the interview.
    Interviewer: Okay, here we go.
    Interviewer: You can sit back.
    Interviewer: Can you sit in your chair?
    Interviewer: How was the puppy feeling?
    Subject: Happy.
    Interviewer: How was the boy feeling?
    Subject: Happy.
    Interviewer: How did you feel when you watched that part?
    Subject: Happy.
    Subject: Ugh.
    clip: You've just got to be kidding me.
    Interviewer: How was the puppy feeling?
    Subject: He was feeling bad.
    Interviewer: How was the boy feeling?
```

```
        Subject: He was feeling, um, angry.
        Interviewer: How did you feel when you watched this part?
        Subject: Bad.
        Subject: Sad.
        Subject: Angry.
        Subject: Boys.
        Subject: Something.
        Subject: Get the boy and the puppy.
        Subject: I miss the boy.
        Interviewer: How's the puppy feeling?
        Subject: Very happy.
        Interviewer: How's the boy feeling?
        Subject: Super happy.
        Interviewer: How did you feel while you were watching that part?
        Subject: Happy.
        Interviewer: That's all, thank you.

    (example 4 output):
        "I hope you enjoyed the last movie. Have you seen it before?" -> "Yeah, The puppy cartoon? Yeah."
        "Can you tell me what happened in the movie? Try to tell the whole story. Remember that stories have a beginning,
things that happen, and an ending." -> "So first there was a boy that was playing a game. So when his mother came in the
door was a box. She put it down and then a puppy was in the box. But the boy didn't want to play. Instead he wanted to
play on his TV. So first he just kicked the puppy, so then he just came back and he did it again. So then he just went
crazy and then got a ball. And then he wanted to play, but then the boy just kicked the ball in the box and the puppy ran
with it. And then he dropped the ball back. And then the boy stopped when he was playing, and then he went to lose the
puppy."
        "Do you remember anything else from the story?" -> "That's all I can remember."
        "What are some of the things you liked about the movie?" -> "When you like. And then all of the puppies ran with
it and then he got trapped in the box."
        "What are some of the things you didn't like about the movie?" -> "When he kicked the puppy."
        "Who gave the boy a box" -> "His mom"
        "What was in the box?" -> "A puppy"
        "What was the boy doing before he got the box?" -> ""Playing on his TV."
        "What was the puppy playing with?" -> "A ball."
        "How are the puppy and the boy the same?" -> "They're both playing with them."
        "In the movie, who is missing a leg? The boy, the puppy, both the boy and the puppy, or no one?" ->  "The boy and
the puppy."
        "How do you think the puppy was feeling? (after watching first clip)" -> "Happy."
        "How do you think the boy was feeling? (after watching first clip)" -> "Happy."
        "And how did you feel while you were watching that part? (after watching first clip)" -> "Happy. Ugh."
        "How do you think the puppy was feeling? (after watching second clip)" -> "He was feeling bad."
        "How do you think the boy was feeling? (after watching second clip)" -> "He was feeling, um, angry."
        "And how did you feel while you were watching that part? (after watching second clip)" -> "Bad."
        "How do you think the puppy was feeling? (after watching third clip)" -> "Sad."
        "How do you think the boy was feeling? (after watching third clip)" -> "Angry."
        "And how did you feel while you were watching that part? (after watching third clip)" -> ""
        "How do you think the puppy was feeling? (after watching fourth clip)" -> "Very happy."
        "How do you think the boy was feeling? (after watching fourth clip)" -> "Super happy."
        "And how did you feel while you were watching that part? (after watching fourth clip)" -> "Happy."
        "Rating" -> "0.8"


### This is the transcript you have to process:
<<<transcript>>>
"""
```