RESEARCH   **Open Access**

Investigating frank autism: clinician initial impressions and autism characteristics

Rebecca R. Canale[1*] , Caroline Larson[2,3] , Rebecca P. Thomas[4,5] , Marianne Barton[1] , Deborah Fein[1] and Inge-Marie Eigsti[1*]

Abstract

Background "Frank autism," recognizable through the first minutes of an interaction, describes a behavioral presentation of a subset of autistic individuals that is closely tied to social communication challenges, and may be linked to so-called "prototypical autism." To date, there is no research on frank autism presentations of autistic adolescents and young adults, nor individuals diagnosed with autism spectrum disorder (ASD) in childhood who do not meet diagnostic criteria during or after adolescence (loss of autism diagnosis, LAD). In addition, there are currently no data on the factors that drive frank autism impressions in these adolescent groups.

Methods This study quantifies initial impressions of autistic characteristics in 24 autistic, 24 LAD and 26 neurotypical (NT) individuals ages 12 to 39 years. Graduate student and expert clinicians completed five-minute impressions, rated confidence in their own impressions, and scored the atypicality of behaviors associated with impressions; impressions were compared with current gold-standard diagnostic outcomes.

Results Overall, clinicians' impressions within the first five minutes generally matched current gold-standard diagnostic status (clinical best estimate), were highly correlated with ADOS-2 CSS, and were driven primarily by prosodic and facial cues. However, this brief observation did not detect autism in all cases. While clinicians noted some subclinical atypicalities in the LAD group, impressions of the LAD and NT groups were similar.

Limitations The brief observations in this study were conducted during clinical research, including some semi-structured assessments. While results suggest overall concordance between initial impressions and diagnoses following more thorough evaluation, findings may not generalize to less structured, informal contexts. In addition, our sample was demographically homogeneous and comprised only speaking autistic participants. They were also unmatched for sex, with more females in the non-autistic group. Future studies should recruit samples that are diverse in demographic variables and ability level to replicate these findings and explore their implications.

Conclusions Results provide insights into the behavioral characteristics that contribute to the diagnosis of adolescents and young adults and may help inform diagnostic decision making in the wake of an increase in the demand for autism evaluations later than childhood. They also substantiate claims of an absence of apparent autistic characteristics in individuals who have lost the diagnosis.

*Correspondence:
Rebecca R. Canale
rebecca.canale@uconn.edu
Inge-Marie Eigsti
inge-marie.eigsti@uconn.edu

Full list of author information is available at the end of the article

Background

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by persistent deficits in social-emotional reciprocity as well as the presence of restricted and repetitive behaviors and interests (RRBI) [1]. In the *Diagnostic and Statistical Manual* (*DSM-V*), ASD is defined as a spectrum of behaviors marked by heterogenous core and co-occurring features [1], diagnosable according to behaviorally defined criteria [2] by trained clinicians. A full diagnostic assessment involves several hours of expert clinician time and standardized assessments; however, studies indicate that nearly all clinicians (97%) report forming an immediate, strong impression of diagnostic status in many cases [3]. Such impressions are generally consistent with gold-standard diagnoses [4]. This impression represents the distinct behavioral presentation of a subset of autistic individuals, dubbed "frank autism," purportedly recognizable within minutes. The current study evaluates the consistency of frank autism impressions in adolescents and young adults, including impressions of a group of individuals diagnosed early in life who no longer display autism symptoms, and compares them to gold-standard diagnostic classifications. We also characterize the behavioral factors that contribute to clinician impressions of autism, to better understand how frank autism impressions relate to enduring, core symptoms, and discuss how the presence of "frank autism" relates to the construct of "prototypical autism" [5]. Note that person-first and identity-first language will both be used in this manuscript to acknowledge the diverse preferences within the autism community [6, 7]; in addition, we limit the use of language consistent with the medical model of autism (e.g., "symptoms" and "deficits") in our discussion of diagnostic criteria.

Autism diagnoses and the role of frank autism impressions

DSM-V ASD diagnostic criteria can differentiate autistic individuals from non-autistic individuals and capture the wide variability within the autism spectrum [8]. Indeed, autism is one of the most reliable diagnoses in the DSM-V [9] with Kappa values from 0.60 to ≥ 0.80 (considered moderate to substantial agreement). A study assessing expert clinician team-based diagnoses reported that some clinicians relied heavily on "feeling autism in the encounter," along with quality of collateral report from parents, to inform diagnostic decisions [10]. However, a study utilizing expert clinician consensus to evaluate the reliability of evaluations performed by community clinicians without ASD-specific expertise found suboptimal agreement on diagnostic status [11]. Of the 87 children and young adults, ages 2–25 years, with community ASD diagnoses, 23% were classified by expert clinician consensus as not autistic, illustrating a discrepancy in diagnostic judgments based on available resources (e.g., time, access to diagnostic tools, expert consultation, etc.) and clinician training.

In one study of frank autism, licensed psychologists watched 10-minute video clips of another clinician administering the Autism Diagnostic Observation Schedule [12] to 42 toddlers who had been flagged with possible ASD during screening [13]. After watching a 10-minute clip, clinicians indicated whether they would refer the child for a comprehensive ASD-specific assessment; two referral decisions were made per child by different clinicians (84 total videos). In this sample, 17 (61%) were referred by one or both clinicians for further ASD assessment and ultimately diagnosed with ASD; seven (25%) were referred by one or both clinicians, but ultimately diagnosed with language delays but not autism; and three (11%) were referred by one or both clinicians, but ultimately found to be typically developing, indicating sensitivity of 0.61 and specificity of 0.82. Of the 57 videos for which neither clinician recommended an ASD-specific evaluation, 11 (39%) were ultimately diagnosed with ASD (i.e., these cases were missed by the observing clinicians). These findings suggest that trained clinicians can identify and distinguish autistic symptoms from characteristics of other developmental delays in toddlers with *some* accuracy, based on 10 min of behavioral observation. However, the high number of false negatives suggests that this information alone is insufficient, at least in the case of toddlers. Furthermore, the determination needed in a clinical evaluation requires not just ruling autism in or out, but also differentiating between autism and other conditions – a significantly more challenging endeavor.

A related study explored the initial impressions of trained clinicians for a sample of 294 children ages 1–4 years who were referred for a diagnostic evaluation after being flagged as at-risk for autism on a brief parent-report screener [4]. After five minutes of interaction during the diagnostic evaluation, clinicians paused and indicated their initial diagnostic impression (ASD or non-ASD) and rated their confidence in this initial impression. Results showed that 238 (81%) initial clinical impressions were concordant with the final diagnosis; the autism cases were judged more accurately than the non-ASD cases, with 86 (92%) of the ASD impressions ultimately receiving an autism diagnosis, consistent with a frank autism phenotype. There was a high false negative

or "missed cases" rate: 49 (24%) cases initially viewed as *not* autism ultimately received an ASD diagnosis; false positive rates were far lower (7%). Clinicians were confident in their initial impressions, particularly for non-autistic cases, with an average confidence rating of 3.74 out of 5. These results highlight the ability of trained clinicians to detect ASD from brief behavioral observation, but underscore that some young autistic children (e.g., 18% in this sample) would be missed by an initial diagnostic impression.

A recent study by the same group [14] further explored what behavioral characteristics informed diagnostic impressions within the first five minutes of interaction with 55 toddlers (mean age = 22.9 months) referred for a developmental evaluation due to parent or pediatrician concerns for autism-related behaviors. Junior (e.g., graduate student) and senior (e.g., PhD level) clinicians were asked to rate their diagnostic impression (autistic or non-autistic), their confidence in this impression, and what behaviors contributed to their impression. Consistent with prior findings, clinicians rated 63% of cases that ultimately received an autism diagnosis as autistic and 100% of cases that did *not* receive an autism diagnosis as non-autistic. Both junior and senior clinicians relied on social reciprocity, nonverbal communication, and eye contact to form accurate initial impressions. Additionally, senior clinicians relied on the child's focus of attention in forming accurate impressions of both autistic and non-autistic children, whereas junior clinicians only relied on this behavior in forming accurate impressions of non-autistic children. These results are the first to explore the behaviors that contribute to diagnostic impressions during brief clinical interactions with young children.

Autism evaluations in adulthood
The prevalence of first-time diagnostic evaluations of adolescents and adults has significantly increased in the past decade, in part because of changes in awareness, diagnostic criteria, and professional practice [15, 16]. The assessment of older individuals provides a unique set of challenges that are not present when assessing young children. Typical diagnostic practice relies heavily on parent or caregiver report of the early developmental history of the individual, which may be difficult to obtain or inaccurately recalled years later [15, 17, 18]; it can also display "telescoping" effects, such that caregivers of individuals who currently display stronger adaptive skills are more likely to recall more strengths and fewer delays in early development [19]. This lack of clear developmental history may force clinicians to rely more heavily on current behavioral observation alone. This, in combination with the evidence that some clinicians rely on less operationalized behavioral observations, by "feeling autism in the encounter" [10], may lead some autistic adults to receive an official diagnosis of autism more readily than others. To date, no studies have explored the behavioral factors that impact clinician impressions of autistic adults.

Autistic characteristics and their impact on impressions
Despite clinical and empirical evidence regarding the good reliability of brief initial impression, the specific factors that contribute to this impression in adolescents and adults are unknown. The initial study proposing frank autism [3] surveyed 151 clinicians with autism-specific expertise about their representation and usage of this construct. Results showed that nearly all (97%) believed that something like frank autism exists, and that they could determine whether an individual fits the phenotype of frank autism in roughly the first ten minutes of interaction or observation. The clinicians who were familiar with the construct estimated that roughly 40% of the ASD population exhibits the frank autism phenotype. Clinicians also reported that the most common specific behaviors associated with this phenotype included impairments or atypicalities in reciprocity, vocal prosody, eye contact, motor mannerisms (such as stereotypies), and gait or posture. These findings highlight factors that may impact initial impressions during ASD diagnostic decision making. To date, no studies have empirically tested the endorsements of these behaviors associated with correct or incorrect frank autism impressions in adults.

Gestures, facial expressions, eye contact, vocal prosody, and social reciprocity have each been implicated as atypical in autistic individuals, and relevant for difficulties with social functioning. Compared to neurotypical peers, autistic individuals produce semantically, pragmatically, and motorically atypical gestures [20–25], as well as atypical facial expressions [26, 27], eye contact [28], and vocal prosody [29, 30]. Together, these characteristics may negatively impact autistic individuals' social interactions and elicit impressions of social awkwardness from naïve observers [26, 27, 29, 31]. Initial impressions for expert clinicians and naïve laypeople may reflect similar processes, despite differences in rater goals (e.g., motivation to engage in future social interaction, versus clinical motivation to arrive at an accurate diagnosis) and the nature of ratings (Likert scales measuring the likeliness that an individual has friends versus binary diagnostic ratings).

In summary, the construct of frank autism is widely assumed in clinical practice and is relevant for initial impressions of behavioral atypicalities in non-clinical settings. As such, it is important to establish which behavioral factors contribute to this impression, as they likely have implications for diagnostic decision making (e.g., who is ultimately diagnosed with autism), as well as clinical management.

Loss of autism diagnosis

Although developmental disorders are typically seen as life-long conditions, a series of studies has identified and characterized a group of individuals who were diagnosed with autism in childhood but who no longer meet DSM-V criteria in adolescence, based on ADOS-2 observations, parent and child symptom report, and clinical best estimate. Estimates suggest that 3–25% of children diagnosed with ASD in early childhood fall into this category [32] by adolescence, although a recent study reported that 37% of toddlers diagnosed with ASD lost the diagnosis by early school age [33]. Our research team has extensively studied these types of individuals [34]. Findings indicate that in early development, the "loss of autism diagnosis" (LAD) group had milder symptoms in the social domain, compared to an age-matched currently autistic group, but equally significant difficulties with communication and repetitive behaviors, including the presence of early language delays. Tests of current functioning indicated that, compared to age- and IQ-matched children with a current autism diagnosis and with neurotypical (NT) children with no history of autism, the LAD group had typical or above-average scores on standardized and experimental assessments of language [35–39], social skills [40, 41], and restricted and repetitive behaviors [42]. To date, no studies have explored frank autism in LAD, and whether these individuals present with subtle or overt frank autism behaviors during initial interactions; findings would help to establish the degree to which these individuals continue to display subtle behavioral characteristics of autism. More broadly, understanding frank autism in LAD may be useful in addressing controversies about the nature of the autism diagnosis [43, 44]. For example, Mottron and colleagues have suggested that developing more constrained diagnostic criteria for autism, informed by strong developmental history data, would facilitate clinical ascertainment and homogeneity of research samples [5, 45].

The current study

The current study had three pre-registered aims (see https://osf.io/5tkrn/?view_only=1f0b6bf70d7d4bab9ebf22da7603e647). Our first aim was to evaluate group (autism, LAD, NT) differences in frank autism impressions made by seven graduate-level (clinical psychology PhD student) and two expert PhD-level clinicians as a predictor of current gold-standard diagnosis in an adolescent and young adult sample. Based on prior studies of LAD and autism, we predicted significantly reduced ASD-like impressions in the LAD and NT groups relative to the autism group, and significant positive correlations between initial impressions of frank autism and ADOS-2 Calibrated Severity Scores (CSS).

Second, drawing on the prior frank autism studies of young children, we asked which behaviors were the most salient contributors to frank autism impressions, by assessing rates of atypicality in gesture, eye contact, motor mannerisms, prosody, facial expressions, attentional focus, and shifting attention (including perseverative thinking and distractibility), social reciprocity, and social initiations. We predicted significantly higher (more atypical) ratings for gesture, eye contact, motor mannerisms, prosody, facial expressions, and social reciprocity. We also predicted that attentional focus and social initiations would be similar across groups, as prior literature typically implicates these more infrequent behaviors (that may be difficult to perceive during a brief encounter) as less consistently associated with a frank autistic presentation.

Third, we hypothesized high overall confidence (e.g., 3 or above on a scale of 1–5) in initial impressions, with higher ratings for NT individuals that had never received an autism diagnosis (based on Wieckowski et al., 2021). We also predicted that the confidence ratings for the LAD group would be significantly lower than both the ASD and NT groups due to possible subclinical social impairments. We hypothesized that higher confidence would be significantly associated with eye contact, motor mannerisms, prosody, and social reciprocity, but not gesture, facial expressions, focus/shifting of attention, or social interactions.

Methods

Participants

This study included participants from a larger study of long-term outcomes in autism. Participants who had completed the Autism Diagnostic Observation Schedule-2 [46] were included in the present study. The sample included currently autistic participants ($n=24$; 7 females), participants with a history of ASD who no longer met diagnostic criteria (LAD; $n=24$; 5 females), and participants with a neurotypical developmental history (NT; $n=26$; 15 females). Participant details are summarized in Table 1. The groups did not differ on age, race/ethnicity, mean household income, verbal skills as measured by Penn Verbal Analogies, or nonverbal skills as measured by Penn Matrix Reasoning.

As expected, the ASD group had higher ADOS-2 scores. The LAD group had marginally higher ADOS-2 CSS scores than the NT group, though the means of both groups fell well below the autism threshold. The ASD and LAD groups had more males than the NT group; no participants identified with a gender other than male or female.

Inclusion criteria were based on the aims of a larger study and thus reflect goals of that project (not discussed here). Criteria thus required: no history of intellectual

Table 1 Participant characteristics

	ASD (n = 24)	LAD (n = 24)	NT (n = 26)	F/χ²	Post-hoc comparison
Age (yrs)	21.22(4.50)	22.72(3.71)	22.74(6.41)	1.02	
M: F *	17:7	19:5	11:15	8.14	ASD = LAD > NT
Race	Native Amer = 0	Native Amer = 0	Native Amer = 0	2.01	
	Asian/Pacific Islander = 1	Asian/Pacific Islander = 0	Asian/Pacific Islander = 1		
	African Amer = 0	African Amer = 0	African Amer = 0		
	White = 22	White = 21	White = 21		
	Multiracial = 1	Multiracial = 1	Multiracial = 0		
	Not reported = 1	Not reported = 1	Not reported = 4		
Ethnicity	Latinx = 1	Latinx = 0	Latinx = 2	2.05	
	Not Latinx = 18	Not Latinx = 19	Not Latinx = 18		
	Not reported = 6	Not reported = 4	Not reported = 6		
Household income ($)	90,833(19,497)	98,611(5,892)	87,333(26,795)	10.65	
Penn Matrix Reasoning	18.85(4.34)	20.03(4.14)	20.32(2.64)	1.04	
Penn Verbal Analogies	7.07(1.94)	7.71(1.44)	7.59(1.84)	1.01	
ADOS-2 CSS***	7.54(1.67)	1.75(0.85)	1.27(0.60)	235.70	ASD>
	6–10	1–3	1–3		LAD > NT

Note Data are presented as M(SD), range, or as count variables. Amer = American. Penn Matrix Reasoning and Penn Verbal Analogy scores represent "efficiency," a composite of accuracy and RT. ADOS-2 CSS = Autism Diagnostic Observation Schedule-2 Calibrated Severity Score. ‡ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

disability, per parent report; current cognitive abilities in the normal range, and scores > 77 on the Vineland Adaptative Behavior Scales-3 [47], a parent-report measure of adaptive functioning; no uncorrected visual or hearing impairments; and no severe psychiatric disorders (e.g., bipolar disorder or schizophrenia). Participants with other, less severe, co-morbid psychiatric disorders such as anxiety, depression, and/or attention deficit/hyperactivity disorder were included in the study. Diagnostic evaluation of such conditions was completed via structured clinical interview and self-report as part of the larger diagnostic battery; further discussion of these data is outside the scope of this paper. Inclusion in the autism group required an ASD diagnosis prior to age five years, documented in a written report by a clinician specializing in autism, as well as the presence of early language delay (first words after age 18 months or first phrases after 24 months). In addition, participants in the autism group had to meet criteria for current ASD based on ADOS-2 scores and best estimate clinical judgement. Inclusion in the LAD group required similar early diagnostic criteria as for the autism group; in addition, participants could exhibit no or minimal current symptoms of ASD, as measured by the ADOS-2 and expert clinical judgement, and had to participate in mainstream educational or occupational environment with no ASD-related accommodations. Inclusion in the NT group required no history of developmental disorder per parent report, no first-degree relatives with an ASD diagnosis, and no or minimal current symptoms of ASD based on the ADOS-2 and expert clinical judgement.

Participants were recruited via their participation in prior studies of ASD, through clinician referrals, posts on social media, flyers distributed at schools and organizations that offer services for autistic individuals and their families, at local schools, libraries, and community centers, a university registry of diverse community members interested in research participation, and by snowball recruitment (e.g., asking participants to nominate other potential participants).

Procedures

Participants completed a comprehensive testing battery to confirm diagnostic status, including the ADOS-2 [46] and a parent interview. ADOS-2 administrations were conducted in person or via a validated online protocol [48], and were recorded for later review. In-person participants completed the standard ADOS-2 Module 4 administration, while online participants completed a modified version that excluded the puzzle task and the break. Autism diagnosis required an ADOS-2 raw score of 8 or greater and expert clinical judgement of autism based on behavioral observation. All ADOS-2 recordings were reviewed by a licensed clinical psychologist with autism expertise to confirm diagnostic status. Participants completed additional measures (including a detailed psychiatric interview) not relevant to the current study.

To measure frank autism impressions, seven graduate students (Clinical Psychology Ph.D. students) and two expert Ph.D.-level clinicians reviewed the recording of the first five minutes of the diagnostic session, comprising discussion of the visit agenda, set up, small talk, and, in some cases, a minute of the first structured ADOS-2 activity (the Tuesday story). The graduate clinicians all established ADOS-2 reliability with a research-reliable licensed psychologist; this group also included a post-doctoral speech-language pathologist fellow with autism experience. The graduate clinicians also conducted the ADOS-2 assessments, though it is important

Canale *et al. Molecular Autism*   (2024) 15:48

Page 6 of 15

to note that they did not complete frank autism impression for any participant for which they conducted the diagnostic study visit. Expert Ph.D.-level clinicians were faculty members with decades of experience in autism assessment and diagnosis. Each ADOS-2 recording was reviewed by two graduate clinicians and one expert clinician, for a total of three raters per recording; clinicians did not watch their own administrations. Graduate clinicians each reviewed 21–23 recordings, and expert clinicians each viewed 37–38 recordings. All clinicians were blind to group status prior to viewing the recordings. After reviewing the five-minute video, clinicians completed a Five-Minute Impressions Form [49]; see Appendix A. The form captured participant details (e.g., date of evaluation, date of review, identity of examining clinician and rater), as well as eight behaviors: gesture, eye contact, motor mannerisms, prosody and vocalizations, facial expressions, attention focus and shifting (including perseverative thinking and distractibility), social reciprocity, and social initiations. Each item was assigned an item-level Score on a 0–2 Likert scale, with "0" representing typical or expected behavior in the category, "1" representing mildly atypical behavior, and "2" representing definitely atypical behavior; this scoring structure is analogous to that used in the ADOS-2. Raters were instructed to respond to all items. If an item did not inform their impression, they were instructed to score that item as 0 and make a note; this score was subsequently converted to a 9, indicating that the item did not contribute to the overall impression, and was not included in the total score calculation. Items were scored based on the rater's observations and their clinical knowledge of typical age-appropriate behavior in that context. Items were summed to form an initial impression total score ranging from 0 (no atypical behaviors detected) to 16 (definitely atypical behavior in eight items). Raters provided initial impressions (autistic or non-autistic) and rated their confidence in this diagnosis from 1 (not very confident) to 5 (extremely confident). Initial impressions from the three raters were averaged, with 0.0 indicating non-autism and 1.0 indicating autism; intermediate scores thus indicated disagreement among the three raters. Confidence scores were also averaged across the three raters.

The Five-Minute Impressions Form was based on Wieckowski et al.,(2021) and Thomas et al., (2024) and modified for use in adolescent and young adult populations via discussions with a large study team including expert clinicians and clinicians in training. A pilot form was employed in evaluations of 10 participants, and further refined via discussion with the study team. Inter-rater reliability scoring for initial impressions of three participants was exceptionally high (Cronbach's alpha=1.0). Training was performed to ensure adequate agreement on the definition of each item-level behavior.

Inter-rater reliability for item-level characteristics was not performed during this initial validation, as we expected variability across individual raters as to which behaviors contributed to their initial impressions.

Measures

Participants completed a battery of measures as part of the larger study, a subset of which were included in the present study analyses. Participants and parents or caregivers completed an online Qualtrics survey probing sociodemographic information of the participant including race and ethnicity, sex assigned at birth, and yearly gross family income.

ADOS-2. The Autism Diagnostic Observation Schedule, Second Edition [46] Module 4, served to confirm diagnosis and to provide a measure of autism-related behavioral characteristics. The ADOS-2 consists of a series of semi-structured tasks designed to elicit social, communicative, and repetitive and stereotyped behaviors relevant to the ASD diagnosis. Module 4, chosen based on developmental level, includes 32 scorable items, scored from 0 (typical) to 3 (definitely atypical). Item scores are used to calculate two Domain Scores (Social Affect and Restricted and Repetitive Behaviors), and to calculate the ADOS-2 Overall Total and Calibrated Severity Score (CSS). A CSS of eight or more suggests an ASD diagnosis. The ADOS-2 CSS is a reliable index of the severity of autism symptoms for each module and yields greater sensitivity and specificity than the ADOS-2 Module 4 raw scores (sensitivity=89.6 (raw scores) and 90.5 (CSS); specificity=72.2 (raw scores) and 82.2 (CSS); [50]. ADOS-2 classifications have good concurrent validity with clinical best estimate of ASD diagnoses [50]. This study used the CSS from administration of the Module 4 revised algorithm, along with information from participant evaluations performed early in development to inform clinical judgment.

Cognitive ability. Participants completed two tasks from the Penn Computerized Neurocognitive Battery (CNB); [51]. The Penn CNB, modeled on standardized neuropsychological tests, provides a reliable online estimate of cognitive functioning in a five-to-eight-minute test. Cronbach's alphas for subtests range from moderate to high (0.78–0.97) with high internal consistency for speed (alpha=0.78–0.98) and moderate internal consistent for accuracy (alpha=0.55–0.95); [52]. Subtests included in this study were the Abbreviated Verbal Reasoning Test, in which participants answer multiple choice questions about verbal analogies, and the Matrix Reasoning Test, in which participants complete visual puzzles. The Abbreviated Verbal Reasoning test has high concordance with the full Penn verbal reasoning battery, R^2=0.90–0.92 [53]. The Matrix Reasoning Test forms part of the Nonverbal Reasoning domain; it was found to

load appropriately in both exploratory and confirmatory bifactor analyses (loading=0.32–0.49); [54]. Following standard procedures [52], we transformed accuracy and RT into an efficiency score, calculated as percent accuracy divided by log RT, to yield individually interpretable scores.

Planned analyses

We used frequentist statistics in R-Studio [55] to test each research question. Significance values were Bonferroni corrected for multiple comparisons. Across analyses, we evaluated 3-group comparisons (group), as well as comparing the autism group to the groups with a non-autism outcome (LAD, NT). First, t-tests were used to assess significant differences in accuracy (e.g., concordance of initial impression with final diagnostic classification of ASD or non-ASD) for graduate versus expert PhD level clinicians. Sensitivity, specificity, positive predictive value, and negative predictive value were calculated and t-tests were used to evaluate differences in accuracy and total score by diagnostic status (e.g. ASD, non-ASD). Analysis of variance models (ANOVA) evaluated group differences (e.g., ASD, LAD, NT) in total score and accuracy, with post-hoc two-way comparisons for any significant results. The relationship between total score and ADOS-2 CSS was assessed using a generalized linear model collapsed across groups.

To evaluate behavioral factors that contribute to initial impression, analysis of variance models were used to compare item-level factors (e.g., gesture, motor mannerisms, eye contact, prosody and vocalizations, facial expressions, focus/shifting of attention, social reciprocity, and social initiations) by group, with post-hoc two-way comparisons for any significant results. Within groups, Pearson correlations assessed the relationship between item-level and initial impression. To assess which item-level best predicted initial impression scores, a generalized linear model was used with each item-level score added as a predictor in the model.

Analysis of variance models were used to compare initial impression confidence across groups, with post-hoc group two-way comparisons for any significant results. There was no missing data across all variables utilized in the proposed analyses.

Results

Preliminary analyses

We compared the impression accuracy of graduate and expert clinicians (79% and 78%, respectively), which did not differ, $t=0.107$, $p=0.92$, $d=0.02$. Total scores also did not differ by expertise, $t=0.447$, $p=0.66$, $d=0.07$. As such, all ratings were collapsed, and mean ratings were used for all subsequent analyses. Similarly, we evaluated accuracy and total score as a function of modality (videoconference, $n=68$, versus in-person, $n=6$). There was no difference by modality for accuracy, $t=0.389$, $p=0.7$, $d=0.22$, or total score, $t=-1.081$, $p=0.32$, $d=0.58$, and session modality was collapsed for all subsequent analyses.

Accuracy of initial impressions

Initial impressions across the three raters were generally concordant with final diagnosis (79% accurate). Of those with an initial impression of ASD ($n=14$), 86% were in the final ASD group, 14% were in the LAD group, and 0% were in the NT group, resulting in an overall accuracy of 86%; Fig. 1. Of those with a non-ASD initial impression ($n=40$), 10% were in the ASD group, 30% were in the LAD group, and 60% were in the NT group, resulting in an overall accuracy of 90% (LAD and NT groups combined). Of those who received initial impressions of both ASD and non-ASD from the three raters (e.g., those for whom clinicians disagreed in initial impressions; $n=20$), 40% were in the ASD group, 50% were in the LAD group, and 10% were in the NT group. Accuracy was greater for non-autistic participants, $t=-7.319$, $p=0.03$, $d=0.63$. The three groups differed in accuracy, $F(2, 71)=6.9$, $p=0.002$. Specifically, initial impressions were significantly more accurate for the NT group compared to the ASD group, $t=-3.582$, $p=0.001$, $d=1.05$, and the LAD group, $t=-3.122$, $p=0.004$, $d=0.23$; see Fig. 2. Accuracy for ASD and LAD groups did not differ, $t=-0.793$, $p=0.43$, $d=0.91$.

Each individual rater's initial impressions were used to calculate sensitivity (the proportion of individuals with a final diagnosis of autism who were judged autistic on initial impression; that is, initial impression true positives divided by final diagnoses of autism), specificity (the proportion of individuals with a final diagnosis of non-autism who were judged non-autistic on initial impression), positive predictive value (PPV; the likelihood that an individual with initial impressions of autism would receive a final diagnosis of autism), negative predictive value (NPV; the likelihood that an individual who received initial impression of non-autism received a final diagnosis of non-autism), false negatives (e.g., the proportion of individuals who received an initial impression of non-autism but ultimately received an autism diagnosis), and false positives (e.g., the proportion of individuals who received an initial impression of autism but ultimately received a diagnosis of non-autism). These calculations were performed to compare the outcome of brief impressions to the outcome of a longer, more thorough clinical diagnostic evaluation.

Results indicated that sensitivity was 66.7%, considered moderately low. Specificity was calculated at 88.0%, considered high. The false positive rate was 34.7%, and the false negative rate was 14.7%. Similarly, PPV was
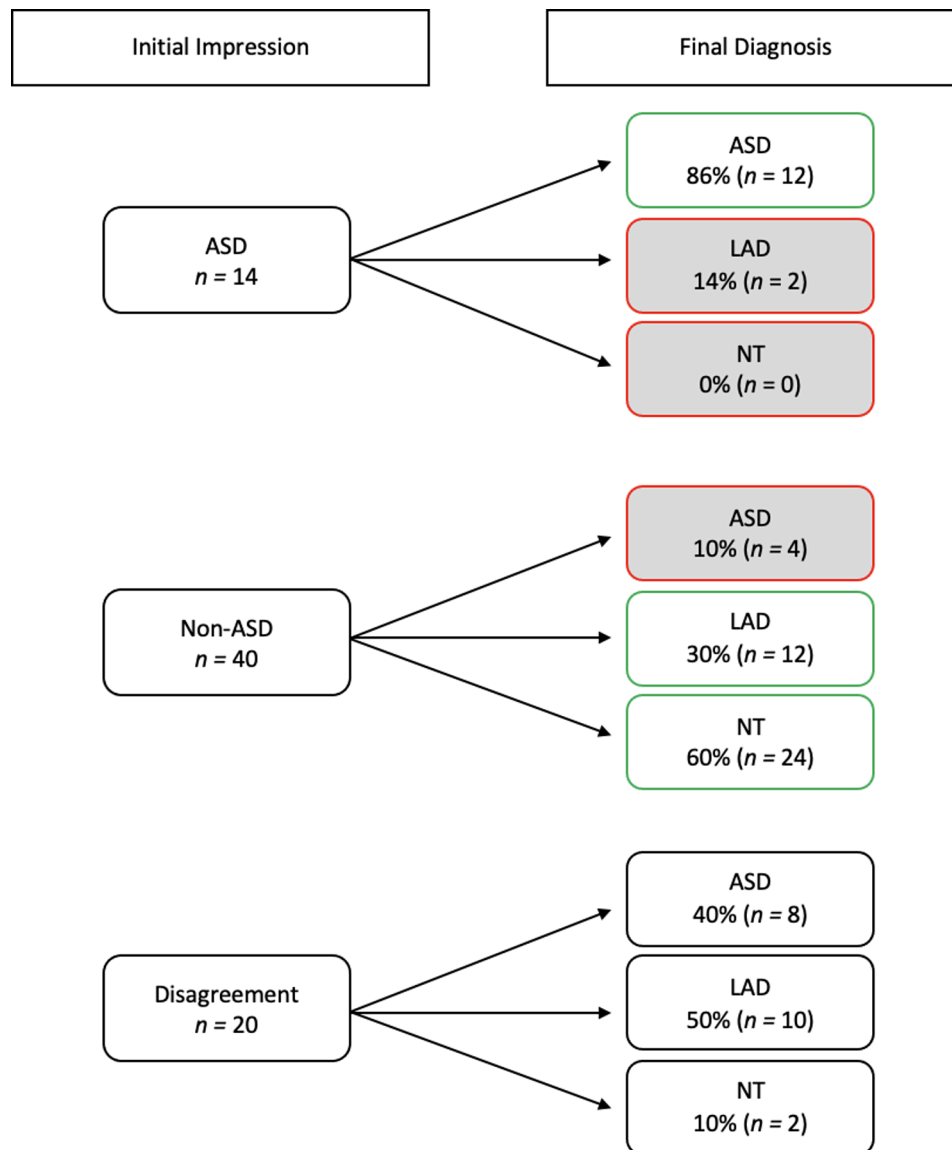
Fig. 1 Accuracy of initial impressions. *Note* Matches between clinician's initial impressions and group are highlighted in green with an unshaded box; mismatches are shown in red with shaded boxes. Disagreement = differences in initial diagnostic impression (ASD versus non-ASD) among the three raters

moderate (72.7%) and NPV was high (84.6%). These findings were broadly consistent with results from Wieckowski et al. (2021), which reported sensitivity = 64% and specificity = 96% (PPV and NPV were not reported).

Total scores
Total scores of atypical behaviors differed by group, $F(2,71) = 44$, $p < 0.001$. Total scores were significantly higher for the ASD group, M(SD) = 4.01(2.02), compared to both the LAD group, M(SD) = 1.11(0.99), $t = 6.319$, $p < 0.001$, $d = 1.82$, and NT group, M(SD) = 0.56(0.85), $t = 7.698$, $p < 0.001$, $d = 2.24$; the LAD and NT groups did not differ, $t = 1.989$, $p = 0.053$, $d = 0.57$; see Fig. 3. Total

scores and ADOS-2 CSS scores were significantly and strongly correlated, $R^2 = 0.65$, $p < 0.001$; see Fig. 4.

Factors contributing to initial impressions
Item-level scores on gesture, eye contact, prosody, facial expressions, social reciprocity, and social initiations differed significantly by group. In all cases, item-level Scores were significantly higher for the ASD group compared to both the LAD and NT, which did not differ; see Table 2. There were no group differences for motor mannerisms or attention.

The relationship between item-level scores and initial impression varied by diagnostic status (autism, non-autism) and group (ASD, LAD, NT). For those in the
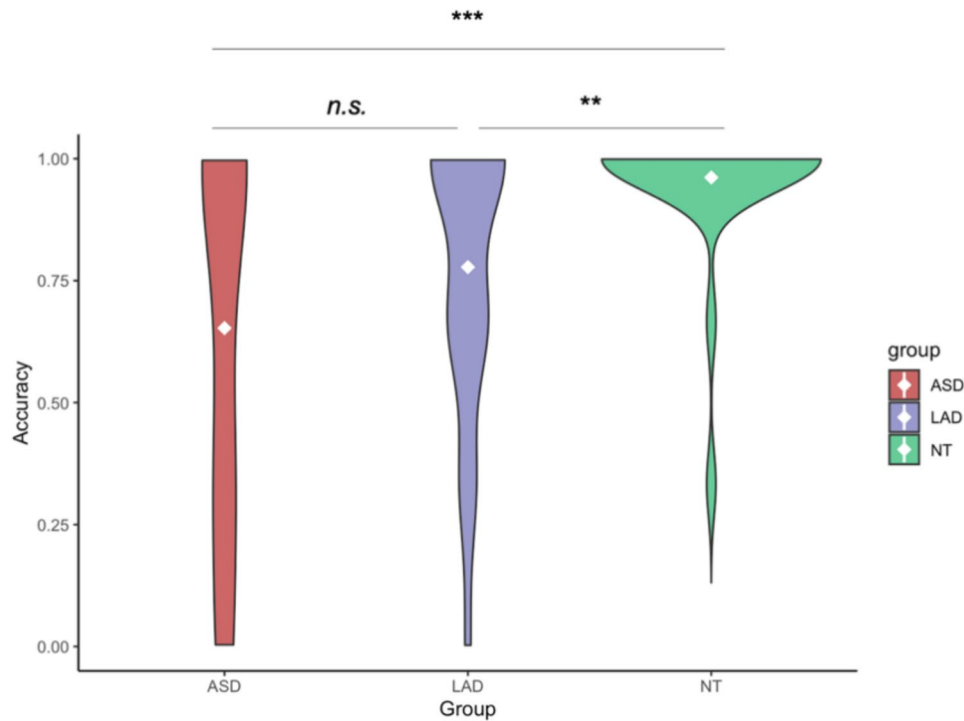
Fig. 2 Initial impression accuracy by group. *Note* Violin plot of initial impression accuracy by group. White diamonds indicate the mean. ** $p < 0.01$, *** $p < 0.001$
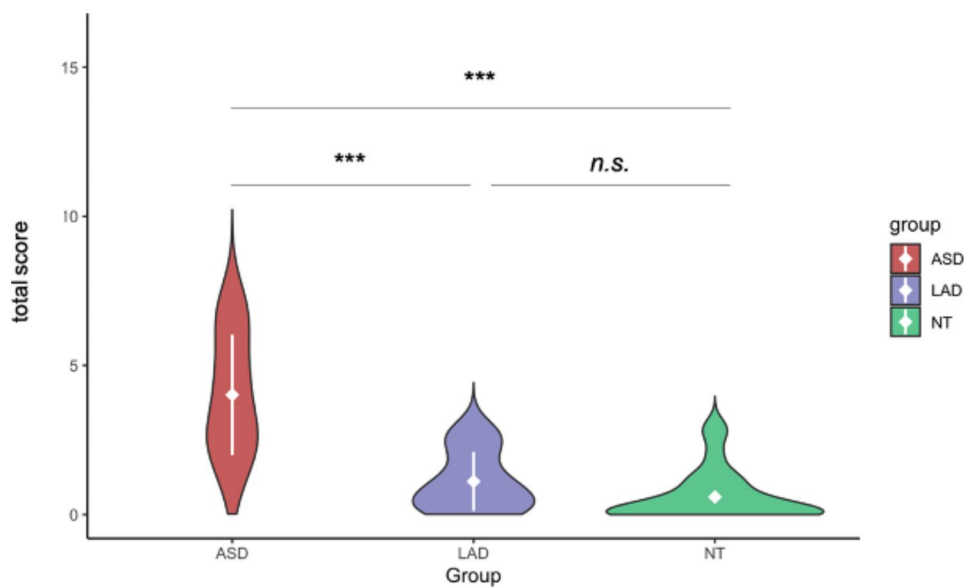

Fig. 3 Total score by group. *Note* White diamonds and lines indicate $M(SD)$. *** $p < 0.001$

ASD group, initial impressions were significantly and strongly correlated with prosody, facial expressions, and social reciprocity. In the LAD group, initial impressions were significantly and strongly correlated with prosody and facial expressions. In the NT group, initial impressions were significantly and strongly correlated with gesture, eye contact, prosody, and facial expressions; see Table 2. Collapsed across group, initial impression scores were most strongly predicted by item-level scores on prosody and facial expressions; Table 3.

Confidence in initial impressions
Confidence was high overall (mean of 3.06 out of 5) and ranged from 2.91 to 3.43, with the highest confidence ratings for the NT group, though rating differences by group were not significant. As noted above, expert and graduate
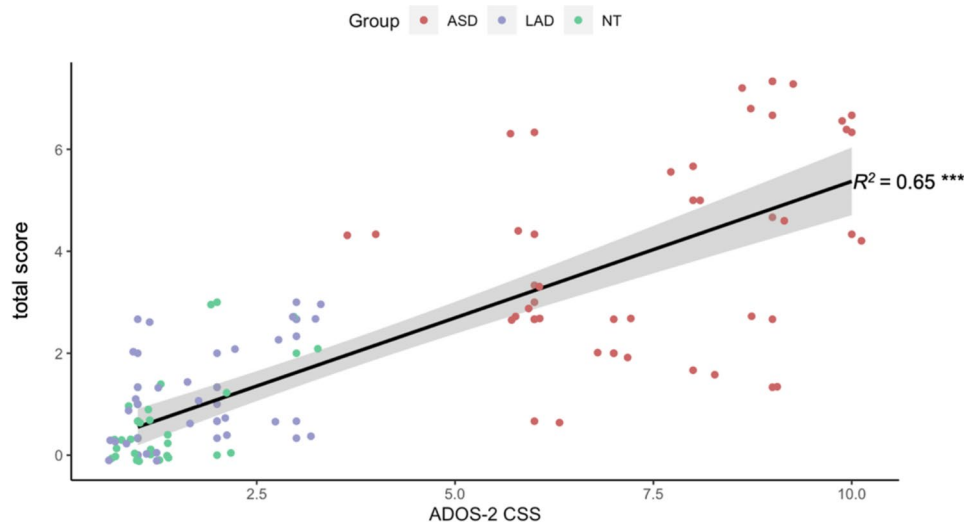
Fig. 4 Relationship between initial impression scores and ADOS-2 CSS scores. *Note* Black line represents the line of best fit; grey shading indicates SE. ***$p < 0.001$

Table 2 Item-level scores by group

	ASD M(SD)	LAD M(SD)	NT M(SD)	F	Post-hoc comparison	d	Correlation (*r*) with initial impression Scores		
							ASD	LAD	NT
Prosody	1.06 (0.63)	0.36 (0.47)	0.06 (0.21)	**29.35***	ASD > NT, LAD	2.13 1.24 0.83	**0.82***	**0.80***	**0.79***
Facial Expressions	0.67 (0.38)	0.21 (0.27)	0.14 (0.21)	**22.85***	ASD > NT, LAD	1.72 1.38 0.27	**0.73***	**0.55***	**0.68***
Gesture	0.60 (0.50)	0.19 (0.33)	0.10 (0.18)	**13.26***	ASD > NT, LAD	1.33 0.96 0.32	0.18	0.15	**0.69***
Eye contact	0.55 (0.42)	0.07 (0.16)	0.06 (0.13)	**25.75***	ASD > NT, LAD	1.57 1.49 0.04	0.30	0.14	**0.56***
Social Reciprocity	0.45 (0.36)	0.13 (0.29)	0.06 (0.16)	**13.34***	ASD > NT, LAD	1.40 0.99 0.26	**0.58***	0.10	-0.11
Social Initiations	0.51 (0.44)	0.14 (0.21)	0.08 (0.18)	**15.06***	ASD > NT, LAD	1.30 1.09 0.28	0.51	0.11	0.13
Attention	0.20 (0.31)	0.05 (0.13)	0.01 (0.07)	**6.24***	No differences	0.85 0.63 0.34	0.31	-0.20	-0.05
Motor Mannerisms	0.17 (0.27)	0.09 (0.24)	0.07 (0.18)	1.19	No differences	0.42 0.30 0.10	0.29	-0.04	0.06

Note Data are represented as M(SD). * Indicates significance after correcting for multiple comparisons of ANOVA (0.05/8; $p < 0.006$), t-tests (0.05/27; $p < 0.002$), and Pearson's correlations (0.05/9; $p < 0.006$). Significant associations are **bolded**. Cohen's d values are listed in the following order: ASD vs. NT, ASD vs. LAD, and LAD vs. NT

clinicians did not differ in confidence. Overall, clinicians were significantly more confident with initial impressions that were correct (confidence = 3.43) compared to those that were incorrect (confidence = 2.83), $t = 3.249$, $p = 0.006$, $d = 0.77$, or in disagreement (confidence = 2.40), $t = 4.969$, $p < 0.001$, $d = 1.30$.

Discussion

The present study evaluated frank autism impressions in adolescents and adults, compared them to current gold-standard diagnostic group classifications, and characterized the behavioral factors that contributed to clinician impressions and confidence in initial impression. Overall,

Table 3 Generalized linear model of item-level scores and initial impression

Predictors	Initial Impression			
	Estimate	Std. Error	t	p
Prosody	0.448	0.044	10.129	< 0.001***
Facial Expressions	0.300	0.087	3.439	0.001**
Social Initiations	0.120	0.084	1.419	0.161
Social Reciprocity	0.032	0.082	0.389	0.698
Eye Contact	0.003	0.079	0.035	0.973
Intercept	-0.003	0.026	-0.138	0.891
Motor Mannerisms	-0.007	0.093	-0.071	0.943
Attention	-0.058	0.102	-0.569	0.571
Gesture	-0.111	0.062	-1.780	0.080

Note All predictors were added to the model simultaneously. AIC = -54.819

the specificity and NPV of initial impressions of a combined group of graduate and expert clinicians with specialized training in autism were high (88.0% and 84.6%, respectively), indicating a low false positive rate; that is, clinicians were highly likely to identify an individual who was non-autistic as such on initial impression. In contrast, the sensitivity and PPV of initial impressions were lower (66.7% and 72.7%, respectively), indicating a **high false negative rate**; about a third of the autistic individuals were misidentified as non-autistic. These results were highly consistent with findings from three previous studies assessing initial impressions of autistic toddlers [4, 13, 14], adding evidence that brief clinical observations provide valuable insight about diagnostic status in *some* but not *all* cases of autism, and are more useful in ruling *out* the presence of autism symptoms than in ruling them *in*. That is, *all* individuals who gave a consistent initial impression of autism also met diagnostic criteria for autism upon full evaluation. In addition, these results are novel in suggesting that clinicians are able to detect autism symptoms after a brief observation of adolescents and adults, not just young children. This may indicate that frank autism presentations may persist through adulthood, though the specific behaviors that contribute to this impression vary over development.

These results represent a novel assessment of whether individuals who no longer present with symptoms of autism (e.g., have lost the ASD diagnosis) present with subclinical, subtle autistic characteristics. The LAD and NT groups did not differ in initial impression, with both groups showing significantly lower scores than the ASD group, suggesting that the LAD group overall presents as neurotypical. However, although misclassifications were relatively infrequent, the accuracy of initial impressions (e.g., alignment with final diagnosis) did differ by group. Raters were significantly more accurate for the NT group compared to both the ASD and LAD groups; clinicians were more likely to have an initial impression of ASD for LAD relative to NT participants. This result suggests that

some individuals with LAD have subtle persistent autistic behaviors.

In line with the sensitivity and specificity values, frankness of initial symptom presentation (initial impression score) was consistent with symptom severity as observed through a lengthy diagnostic observation (ADOS-2 CSS). Importantly, in this study, impression ratings were provided by clinicians who did not perform the diagnostic evaluation, suggesting that the identification of frank autism is not a result of confirmation bias (i.e., a tendency for clinicians to align their final diagnosis with their initial impressions).

Autistic characteristics and initial impressions

To expand our understanding of how specific behaviors contribute to initial impressions of frank autism in adolescents and adults, we assessed eight behaviors that have been described as most central to impressions in anecdotal [3] and empirical [4, 13, 14] research. Results indicated that ratings within the first five minutes on prosody and facial expressions were the best predictors of initial impression across final diagnosis (autistic, non-autistic) and group (ASD, LAD, NT). That is, atypical prosody and facial expressions appear to be the most salient and reliably available indicators of ASD diagnoses in brief observations of adults; if these behaviors are seen as typical, the individual is more likely to receive an initial impression and a final determination as non-autistic. Because of the low false positive rate in this relatively small sample, further analyses were not performed to determine which behaviors contributed to group differences in accuracy; this should be examined in future research. Prosody and facial expressions were significantly correlated with initial impression scores when evaluated within each of three groups. In addition, for the ASD group, social reciprocity was a strong predictor of impressions, and for the NT group, gestures and eye contact were strong predictors. This indicates that, while prosody and facial expressions are the most prominent behavioral features of frank autism impressions in adults, clinicians rely on other behavioral factors as well.

This finding was also reflected in clinician confidence in initial impressions, in that higher scores on gesture, eye contact, prosody, facial expressions, and social reciprocity were all significantly negatively correlated with confidence for the non-ASD diagnostic status (LAD and NT). That is, if a clinician formed an initial impression of non-autism, but observed mild atypicality in one or more of behavioral domains, they might still settle on a non-ASD impression, but with reduced confidence. Even though overall confidence did not vary by group, the absence of significant relationships between behavioral factors and confidence ratings for the autism group may indicate that confidence in initial impressions is more susceptible to

change based on contradictory evidence (i.e., presence of atypical behaviors for a generally non-autism impression) in LAD and NT groups as compared to contradictory evidence (i.e., absence of atypical behaviors for a generally autistic impression) in the ASD group; this is consistent with the broad finding that the presence of evidence is more salient than the absence of evidence [56].

Clinical implications

One clinical implication of the current study is that *brief observations alone are not sufficient to detect all cases of autism accurately.* Expert and graduate clinicians with specialized autism-specific training did not have an initial impression of autism for roughly 33% of autistic individuals. *Longer structured assessments by trained clinicians provide invaluable information about the nature and severity of autism symptoms.* In practice, clinician judgment should be considered an integral, but not the sole, factor in diagnosis.

The current study also provides important information about what behaviors inform diagnosis of adolescents and adults. Typical diagnostic practices rely heavily on parent or caregiver report of the developmental history of the individual, which may be difficult to obtain and less accurately recalled years later [15, 17, 18]. The absence of a clear developmental history may force clinicians to rely more heavily on behavioral observations alone. This, in combination with the evidence that some clinicians rely on more abstract behavioral observations, may lead some autistic adults to receive an official diagnosis of autism more readily than others. For example, if prosody and facial expressions are the most salient diagnostic cues, an individual who presents with less frankly autistic behaviors in those domains may receive an initial impression of being non-autistic. When combined with a limited developmental history, such a presentation may lead to an incorrect non-ASD Impression. This factor may also contribute to under-diagnosis of autistic females [57].

Implications for our understanding of autism as a diagnosis

There has been a resurgence of the debate about the loosening of diagnostic criteria for autism in the DSM-V, and the resulting increased prevalence and heterogeneity of the diagnosis [43, 45]. While these DSM changes present challenges for how best to identify and diagnose participants for autism research, we note that the *clinical* question of how to diagnose autism requires, by definition and practice, a reliance on current DSM or ICD criteria, informed by a detailed semi-structured clinical interview in combination with a thorough developmental history. In research, Mottron and colleagues have proposed that cohorts of autistic individuals should be selected based on the "prototypicality" of their autism

symptoms, as judged by expert clinicians [45]. This manuscript asks how clinicians' initial diagnostic impressions ("five-minute impressions" in the current study) correspond to a full diagnostic impression; the current study also tests whether brief five-minute impressions provide additional information about the clinical presentation of autism in individuals who have lost the autism diagnosis. As reported by De Marchena and Miller, 2017 [3], many expert clinicians report a strong initial impression of frank autism; it is *clinically* and *diagnostically* relevant to contrast these initial impressions with the findings of a full diagnostic evaluation, informed by clinical history and by a semi-structured interview, as one's initial impressions cannot help but inform one's subsequent evaluation. The current results indicate that initial impressions of frank autism, which likely overlaps with the construct of "prototypical" autism, are highly sensitive; clinicians rarely "felt" autism that turned out *not* to be autism on full evaluation. However, results also suggested that some cases that did not yield an initial frank or prototypical impression of autism during a brief five-minute interaction, were judged to be autism on full evaluation. This is not surprising, particularly in a population with strong verbal and cognitive abilities. These results suggest that, even for adult individuals who present as frankly autistic, the behavioral factors that contribute to this impression vary, encompassing prosody, facial expressions, and other behaviors. Comparing behaviors used for initial impressions in older adolescents and adults versus those used for toddlers [14] also suggests that quite different behaviors are used for different ages or functioning levels, in part because young children may not have enough language to judge prosody or other more mature behaviors.

The current study established a strong association between initial impression and ADOS-2 CSS, indicating that individuals who present as frankly autistic may have more apparent or severe autism symptoms. However, those who present as less frankly autistic during initial impressions may still meet clinical criteria when observed through a full diagnostic assessment. This evidence suggests that the ADOS-2 or other structured observation plays a critical role in eliciting behaviors that are core to the autism phenotype that may or may not be present within the first few minutes of observation or interaction. Rather than focusing only on frank autism characteristics, clinical researchers must carefully evaluate individual differences.

There has been ongoing debate about the stability of autism as a lifelong condition. Several research groups have documented individuals who met criteria for autism in childhood but no longer displayed clinical levels of ASD later in development [33, 34, 58–60]. This subset of the autism spectrum may range from 3 to 37%

of children diagnosed with ASD in early childhood [32, 33]. The effect of the transition to independence in late adolescence and adulthood have yet to be characterized in this group of individuals. The current study indicates that, overall, LAD individuals are indistinguishable from their NT peers on initial impression, but that some of them present with subtle subclinical autistic features that are recognizable to trained clinicians and that contribute to reduced accuracy of initial impressions. Importantly, these features do not give rise to clinical level impairments in social functioning, as evidenced by non-autistic range ADOS-2 CSS scores but may have subtle implications for daily functioning that have yet to be fully explored.

Implications for social functioning and quality of life of autistic adolescents and adults
Previous studies have established that naïve raters are sensitive to atypical behaviors, including prosody and facial expressions, that negatively impact initial impressions of autistic individuals and that may lead to a decrease in the rater's willingness to engage with the autistic individual in hypothetical circumstances, such as sharing a meal [26, 27, 31]. The current study provided evidence that clinicians utilize similar cues in clinical diagnosis, indicating that prosody and facial expressions are highly salient for both clinical and social initial impressions. These results have important implications when assessing the social functioning of autistic adolescents and adults. If naïve individuals are less willing to interact with their autistic peers due to an increased perception of social awkwardness, this may exacerbate the social difficulties of autistic individuals, which may in turn impact their quality of life and vulnerability to anxiety or depression. Future studies should explore relationships among initial impressions and other functional domains to better understand how we can best support autistic individuals, as well as provide further evidence to support societal acceptance of neurodiversity.

Limitations
The brief observations in this study were conducted in a clinical research setting, including some portion of structured testing. While results suggest concordance with naïve impressions during typical social interactions, findings may not generalize to less structured, informal contexts. In addition, the content of interactions varied, with some interactions comprising mainly informal "chit chat" during equipment setup, and others comprising a mix of structured and unstructured activities. The ADOS-2 activity at the beginning of the recording was not consistent across participants; in-person participants began with a puzzle task (Construction activity), whereas online participants began with a story activity (Tuesday

Story). Although results revealed no modality-specific differences, this variability could have impacted interactions on a case-by-case basis. That said, the consistency of results across modalities is a testament to the stability and potential clinical utility of initial impressions, given that they seem to generalize to a range of settings. Our sample was relatively small, demographically homogeneous, and comprised of only speaking autistic participants; furthermore, there were significantly more females in the NT group compared to ASD or LAD groups. We were underpowered to compare which behaviors led to disagreement among raters, and to test gender difference, precluding the discovery of which factors played a critical role in the formation of accurate initial impressions. Future studies should recruit larger sample sizes that are diverse across demographic variables and ability level to thoroughly explore these nuances.

Conclusions
The results of the current study indicate that, while clinicians' initial impressions made within the first five minutes of observation of a diagnostic evaluation generally matched current gold-standard diagnostic status and were highly correlated with ADOS-2 CSS, this brief observation was not sufficient to detect autism in all cases. Clinicians relied heavily on atypical prosody and facial expressions when forming an initial impression of autism, indicating that these cues are extremely salient even within a brief observation. Lastly, the results of the current study further established that LAD individuals no longer exhibit clinically significant autism symptoms, but that some individuals in this group may continue to display subtle autism characteristics that lead to more variable initial impressions. Future research is needed to explore the impact of sex and gender on initial impression, as well as the impact of frank autism presentations on other domains of functioning, such as social functioning in everyday life (e.g., making and maintaining friendships and romantic relationships), quality of life, life satisfaction, and comorbid psychopathology.

amended for use in this study. All analyses were performed by RRC with consultation from CL. Funding acquisition was led by IME and DF.

Declarations

Ethics approval and consent to participate
The experimental protocol was approved by the University of Connecticut IRB.

Consent for publication
Not applicable.

Competing interests
Dr. Fein and Dr. Barton are co-owners of M-CHAT LLC, which licenses use of the M-CHAT-R in electronic products. The other authors declare that they have no conflict of interest.

Author details
¹Department of Psychological Sciences, University of Connecticut, 406 Babbidge Road, Storrs, CT 06269, USA
²Department of Speech, Language, and Hearing Sciences, University of Missouri, Columbia, MO, USA
³Interdisciplinary Neuroscience Program, University of Missouri, Columbia, MO, USA
⁴Division of Developmental Medicine, Boston Children's Hospital, Boston, MA, USA
⁵Department of Psychiatry, Harvard Medical School, Boston, MA, USA

References
1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. 2013.
2. Volkmar FR, McPartland JC. From Kanner to DSM-5: Autism as an Evolving Diagnostic Concept. Annu Rev Clin Psychol. 2014;10(1):193–212.
3. De Marchena A, Miller J. Frank presentations as a novel research construct and element of diagnostic decision-making in autism spectrum disorder. Autism Res. 2017;10(4):653–62.
4. Wieckowski AT, De Marchena A, Algur Y, Nichols L, Fernandes S, Thomas RP, et al. The first five minutes: initial impressions during autism spectrum disorder diagnostic evaluations in young children. Autism Res. 2021;14(9):1923–34.
5. Mottron L, Gagnon D. Prototypical autism: New diagnostic criteria and asymmetrical bifurcation model. Acta Psychol (Amst). 2023;237:103938.
6. Anderson-Chavarria M. The autism predicament: models of autism and their impact on autistic identity. Disabil Soc. 2021;37:1321–41.
7. Kenny L, Hattersley C, Molins B, Buckley C, Povey C, Pellicano E. Which terms should be used to describe autism? Perspectives from the UK autism community. Autism Int J Res Pract. 2016;20(4):442–62.
8. Frazier TW, Youngstrom EA, Speer L, Embacher R, Law P, Constantino J, et al. Validation of proposed DSM-5 criteria for Autism Spectrum Disorder. J Am Acad Child Adolesc Psychiatry. 2012;51(1):28–e403.
9. Rice CE, Carpenter LA, Morrier MJ, Lord C, DiRienzo M, Boan A, et al. Defining in Detail and evaluating reliability of DSM-5 criteria for Autism Spectrum disorder (ASD) among children. J Autism Dev Disord. 2022;52(12):5308–20.
10. Hayes J, McCabe R, Ford T, Russell G. Drawing a line in the sand: affect and testimony in autism assessment teams in the UK. Sociol Health Illn. 2020;42(4):825–43.
11. Hausman-Kedem M, Kosofsky BE, Ross G, Yohay K, Forrest E, Dennin MH, et al. Accuracy of reported community diagnosis of Autism Spectrum Disorder. J Psychopathol Behav Assess. 2018;40(3):367–75.
12. Lord C, Rutter M, DiLavore PC, Risi S. Autism Diagnostic Observation schedule (ADOS). Los Angeles: Western Psychological Services; 2002.
13. Gabrielsen TP, Farley M, Speer L, Villalobos M, Baker CN, Miller J. Identifying autism in a brief Observation. Pediatrics. 2015;135(2):e330–8.
14. Thomas RP, De Marchena A, Wieckowski AT, Stahmer A, Milan S, Burke JD, et al. Accuracy of initial diagnostic impressions of autism in toddlers and behaviors that inform these impressions. Autism Res. 2024;17(3):568–83.
15. Huang Y, Arnold SR, Foley KR, Trollor JN. Diagnosis of autism in adulthood: a scoping review. Autism. 2020;24(6):1311–27.
16. Jensen CM, Steinhausen HC, Lauritsen MB. Time trends over 16 years in incidence-rates of Autism Spectrum Disorders across the Lifespan Based on Nationwide Danish Register Data. J Autism Dev Disord. 2014;44(8):1808–18.
17. Lai MC, Baron-Cohen S. Identifying the lost generation of adults with autism spectrum conditions. Lancet Psychiatry. 2015;2(11):1013–27.
18. Rutherford M, McKenzie K, Forsyth K, McCartney D, O'Hare A, McClure I, et al. Why are they waiting? Exploring professional perspectives and developing solutions to delayed diagnosis of autism spectrum disorder in adults and children. Res Autism Spectr Disord. 2016;31:53–65.
19. Hus V, Taylor A, Lord C. Telescoping of caregiver report on the Autism Diagnostic interview - revised: telescoping on the ADI-R. J Child Psychol Psychiatry. 2011;52(7):753–60.
20. De Marchena A, Kim ES, Bagdasarov A, Parish-Morris J, Maddox BB, Brodkin ES, et al. Atypicalities of gesture form and function in autistic adults. J Autism Dev Disord. 2019;49(4):1438–54.
21. De Marchena A, Eigsti I. Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. Autism Res. 2010;3(6):311–22.
22. Morett LM, O'Hearn K, Luna B, Ghuman AS. Altered gesture and Speech Production in ASD detract from In-Person communicative quality. J Autism Dev Disord. 2016;46(3):998–1012.
23. Silverman LB, Bennetto L, Campana E, Tanenhaus MK. Speech-and-gesture integration in high functioning autism. Cognition. 2010;115(3):380–93.
24. Taverna EC, Huedo-Medina TB, Fein DA, Eigsti IM. The interaction of fine motor, gesture, and structural language skills: the case of autism spectrum disorder. Res Autism Spectr Disord. 2021;86:101824.
25. Trujillo JP, Özyürek A, Kan CC, Sheftel-Simonova I, Bekkering H. Differences in the production and perception of communicative kinematics in autism. Autism Res. 2021;14(12):2640–53.
26. Grossman RB. Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. Autism. 2015;19(5):580–7.
27. Sasson NJ, Faso DJ, Nugent J, Lovell S, Kennedy DP, Grossman RB. Neurotypical peers are less willing to interact with those with autism based on Thin slice judgments. Sci Rep. 2017;7(1):40700.
28. Neumann D, Spezio ML, Piven J, Adolphs R. Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. Soc Cogn Affect Neurosci. 2006;1(3):194–202.
29. Weed E, Fusaroli R, Simmons E, Eigsti IM. Different in different ways: A network-analysis approach to voice and prosody in Autism Spectrum Disorder. Lang Learn Dev. 2024;20(1):40-57
30. Grossman RB, Bemis RH, Plesa Skwerer D, Tager-Flusberg H. Lexical and Affective Prosody in Children with High-Functioning Autism. J Speech Lang Hear Res. 2010;53(3):778–93.
31. Sasson NJ, Morrison KE. First impressions of adults with autism improve with diagnostic disclosure and increased autism knowledge of peers. Autism. 2019;23(1):50–9.
32. Helt M, Kelley E, Kinsbourne M, Pandey J, Boorstein H, Herbert M, et al. Can children with Autism recover? If so, how? Neuropsychol Rev. 2008;18(4):339–66.
33. Harstad E, Hanson E, Brewster SJ, DePillis R, Milliken AL, Aberbach G, et al. Persistence of Autism Spectrum Disorder from Early Childhood through School Age. JAMA Pediatr. 2023;177(11):1197.
34. Fein D, Barton M, Eigsti I, Kelley E, Naigles L, Schultz RT, et al. Optimal outcome in individuals with a history of autism. J Child Psychol Psychiatry. 2013;54(2):195–205.
35. Tyson K, Kelley E, Fein D, Orinstein A, Troyb E, Barton M, et al. Language and Verbal Memory in individuals with a history of Autism Spectrum disorders who have achieved optimal outcomes. J Autism Dev Disord. 2014;44(3):648–63.
36. Canfield AR, Eigsti IM, De Marchena A, Fein D. Story Goodness in adolescents with Autism Spectrum disorder (ASD) and in optimal outcomes from ASD. J Speech Lang Hear Res. 2016;59(3):533–45.
37. Suh J, Eigsti IM, Naigles L, Barton M, Kelley E, Fein D. Narrative performance of Optimal Outcome Children and adolescents with a history of an Autism Spectrum disorder (ASD). J Autism Dev Disord. 2014;44(7):1681–94.

38. Irvine CA, Eigsti IM, Fein DA. Uh, um, and Autism: Filler disfluencies as pragmatic markers in adolescents with optimal outcomes from Autism Spectrum Disorder. J Autism Dev Disord. 2016;46(3):1061–70.

39. Fitch A, Fein DA, Eigsti IM. Detail and Gestalt Focus in Individuals with optimal outcomes from Autism Spectrum disorders. J Autism Dev Disord. 2015;45(6):1887–96.

40. Orinstein AJ, Suh J, Porter K, De Yoe KA, Tyson KE, Troyb E, et al. Social function and communication in Optimal Outcome Children and adolescents with an Autism history on structured test measures. J Autism Dev Disord. 2015;45(8):2443–63.

41. Suh J, Orinstein A, Barton M, Chen CM, Eigsti IM, Ramirez-Esparza N, et al. Ratings of broader Autism phenotype and personality traits in optimal outcomes from Autism Spectrum Disorder. J Autism Dev Disord. 2016;46(11):3505–18.

42. Troyb E, Orinstein A, Tyson K, Eigsti IM, Naigles L, Fein D. Restricted and repetitive behaviors in individuals with a history of ASDs who have achieved optimal outcomes. J Autism Dev Disord. 2014;44(12):3168–84.

43. Fein D, Eigsti I, Barton M. Response to a radical change in our autism research strategy is needed: back to prototypes by Mottron et al. (2021). Autism Res. 2021;14(10):2237–8.

44. Eigsti I, Fein D, Larson C. Editorial perspective: another look at 'optimal outcome' in autism spectrum disorder. J Child Psychol Psychiatry. 2023;64(2):332–4.

45. Mottron L. A radical change in our autism research strategy is needed: back to prototypes. Autism Res. 2021;14(10):2213–20.

46. Lord C, Luyster J, Gotham K, Guthrie W. Autism Diagnostic Observation schedule, Second Edition (ADOS-2). Torrance, CA: Western Psychological Services; 2012.

47. Sparrow SS, Cicchetti DV. The Vineland Adaptive Behavior scales. C.S. Newmark, editor; 1989.

48. Eigsti I, Thomas RP, Stabile M, Mohan A, Dieckhaus MFS, Crutcher J, et al. Online administration of the ADOS for research with adolescents and adults in response to the pandemic. Autism Res. 2022;15(10):1909–16.

49. Thomas R, Canale R, Larson C. Five-Minute Initial Impressions Form. 2021.

50. Hus V, Lord C. The Autism Diagnostic Observation schedule, Module 4: revised algorithm and standardized severity scores. J Autism Dev Disord. 2014;44(8):1996–2012.

51. Gur R. Computerized neurocognitive scanning: I. Methodology and Validation in Healthy people. Neuropsychopharmacology. 2001;25(5):766–76.

52. Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, et al. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. J Neurosci Methods. 2010;187(2):254–62.

53. Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC. Development of abbreviated nine-item forms of the raven's Standard Progressive matrices Test. Assessment. 2012;19(3):354–69.

54. Moore TM, Reise SP, Gur RE, Hakonarson H, Gur RC. Psychometric properties of the Penn Computerized Neurocognitive Battery. Neuropsychology. 2015;29(2):235–46.

55. RStudio Team. RStudio: Integrated Development for R. Boston, MA: PBC; 2020.

56. Altman DG, Bland JM. Statistics notes: absence of evidence is not evidence of absence. BMJ. 1995;311(7003):485–485.

57. Cola ML, Plate S, Yankowitz L, Petrulla V, Bateman L, Zampella CJ, et al. Sex differences in the first impressions made by girls and boys with autism. Mol Autism. 2020;11(1):49.

58. Anderson DK, Liang JW, Lord C. Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. J Child Psychol Psychiatry. 2014;55(5):485–94.

59. Moulton E, Barton M, Robins DL, Abrams DN, Fein D. Early characteristics of children with ASD who demonstrate optimal Progress between Age two and four. J Autism Dev Disord. 2016;46(6):2160–73.

60. Fountain C, Winter AS, Bearman PS. Six Developmental trajectories Characterize Children with Autism. Pediatrics. 2012;129(5):e1112–20.

Publisher's note