*Original Article*

autism

Autism
1–10

DOI: 10.1177/1362361316677957
journals.sagepub.com/home/aut
SAGE

# What's the story? A computational analysis of narrative competence in autism

**Michelle Lee[1], Gary E Martin[2], Abigail Hogan[1], Deanna Hano[1], Peter C Gordon[3] and Molly Losh[1]**

**Abstract**

Individuals with autism spectrum disorder demonstrate narrative (i.e. storytelling) difficulties which can significantly impact their ability to form and maintain social relationships. However, existing research has not comprehensively documented these impairments in more open-ended, emotionally evocative situations common to daily interactions. Computational linguistic measures offer a promising complement to traditional hand-coding methods of narrative analysis and in this study were applied together with hand coding of narratives elicited with emotionally salient scenes from the Thematic Apperception Test. In total, 19 individuals with autism spectrum disorder and 14 typically developing controls were asked to tell stories about six images from the Thematic Apperception Test. Both structural and qualitative aspects of narrative were assessed using a hand-coding system and Latent Semantic Analysis, an automated computational measure of semantic similarity. Individuals with autism spectrum disorder demonstrated significant difficulties with the use of complex syntax to integrate their narratives and problems explaining characters' intentions. These and other key narrative skills were strongly related to narrative competence scores derived from Latent Semantic Analysis, which also distinguished the autism spectrum disorder group from controls. Together, results underscore key narrative impairments in autism spectrum disorder and support the promise of Latent Semantic Analysis as a valuable tool for the quantitative assessment of complex language abilities.

**Keywords**

autism spectrum disorder, communication and language, narrative

Narrative (i.e. storytelling) is ubiquitous, emerging early in development and serving as a fundamental form of communication throughout the life span. Across different cultures and languages, narrative is used as a tool for organizing and sharing meaningful experiences with others by imposing temporal and causal order to events and relating them from a psychological stance (Berman and Slobin, 1994; Bruner, 2004; Ochs and Capps, 2001). Narrative impairments are a central feature of autism spectrum disorder (ASD); individuals with ASD narrate less in conversation and when they do, their stories are less coherent and lack integration of protagonists' thoughts and emotions (e.g. Capps et al., 2000; Losh and Capps, 2003; Loveland and Tunali, 1993). Such differences impose serious barriers to successful social communication; thus, careful characterization of these skills in ASD is paramount to better understanding the social profile of this disorder and intervention planning.

Historically, studies of narrative have relied on detailed hand-coding methods. These methods have provided invaluable insights into specific narrative devices proving most problematic in ASD. For instance, analyses of narratives across different chronological and mental ages in ASD have repeatedly detected problems explaining protagonists' actions in relationship to their psychological states, which results in narratives bereft in social and psychological significance (Capps et al., 2000; Colle et al., 2008; Losh and Capps, 2003; Tager-Flusberg and Sullivan, 1995). While these coding schemes have been essential in identifying key narrative deficits in ASD, they are highly

[1]Northwestern University, USA
[2]St. John's University, USA
[3]University of North Carolina at Chapel Hill, USA

**Corresponding author:**
Molly Losh, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA.
Email: m-losh@northwestern.edu

labor-intensive and require extensive training for valid and reliable coding. Computational linguistic measures such as Latent Semantic Analysis (LSA; Landauer and Dumais, 1997) have the potential to address such limitations. LSA generates a quantitative measure of the shared semantic meaning between bodies of text. Building on search techniques developed for document retrieval (i.e. web search), LSA has since been applied to a number of linguistic contexts, including quantifying narrative quality in clinical populations (e.g. Elvevag et al., 2007). Demonstrating the utility of LSA as a sensitive measure of narrative ability in ASD, Losh and Gordon (2014) found that individuals with ASD differed from controls on LSA metrics during narrative recall tasks, but not in a structured storybook task, aligning with prior findings using hand coding to characterize narrative ability (Losh and Capps, 2003). LSA measures also correlated positively with the use of complex syntax and evaluative devices that enrich narratives with a psychological perspective (e.g. explaining characters' thoughts and emotions), suggesting that LSA can capture key aspects of narration impacted in ASD. The potential to capture such differences rapidly and objectively highlights the possible value of LSA in clinical contexts and large-scale research studies, where detailed hand coding may not be feasible.

Furthermore, prior research characterizing narrative abilities in ASD has primarily utilized highly structured tasks such as wordless picture books, which bear little resemblance to the narrative demands common to everyday social interactions. These tasks also substantially reduce the burden on the narrator by offering a temporal series of illustrations that require little interpretation to produce an adequate narrative. Implementing more open-ended stimuli to elicit narratives is therefore an important step in characterizing more precisely the strengths and weaknesses in narrative ability exhibited by individuals with ASD. This study applied such strategies using both hand coding and LSA to examine narratives told by individuals with ASD in response to illustrations from the Thematic Apperception Test (TAT; Murray, 1943). Traditionally used to evaluate psychopathology, the unstructured, emotionally ambiguous nature of TAT images may more closely resemble the demands of narration in everyday interactions. Therefore, this study provides a novel opportunity to comprehensively explore narration in ASD using multiple methods, with key implications for further defining the social phenotype of ASD and clinical intervention.

## Narrative ability in ASD

Previous studies comparing narratives in ASD to typically developing (TD) individuals indicate a particular pattern of strengths and weaknesses. For example, individuals with ASD do not differ from language-matched controls on global aspects of narrative, such as length or identification of basic narrative features (e.g. main characters, setting; Beaumont and Newcombe, 2006; Capps et al., 2000; Hogan-Brown et al., 2013; Losh and Capps, 2003; Tager-Flusberg and Sullivan, 1995). However, individuals with ASD encounter difficulties integrating story elements into a cohesive whole, driven in part by limited use of complex syntax that results in a lack of temporality and connection of plot points (Capps et al., 2000; Losh and Capps, 2003). Additionally, individuals with ASD show limited use of evaluative devices to convey their perspective and infuse story events with broader meaning, instead simply labeling or describing behavioral indices of emotions (Capps et al., 2000; Colle et al., 2008; Losh and Capps, 2003; Tager-Flusberg and Sullivan, 1995). Finally, individuals with ASD more often make pedantic or inappropriate comments during narration, detracting from overall narrative quality (Colle et al., 2008; Losh and Capps, 2003).

Contextual factors appear to be quite important to consider when evaluating the narrative abilities of individuals with ASD. Whereas TD individuals tend to tell more elaborate and coherent narratives in open-ended contexts, individuals with ASD show the greatest difficulties during these unstructured contexts. For instance, Losh and Capps (2003) found that individuals with ASD used less complex syntax, fewer evaluative devices and required significantly more experimenter prompting for clarification when constructing personal narratives during semi-structured conversation, but not when narrating from a more structured wordless picture book. When asked to tell stories about their own past experiences, individuals with ASD also show difficulty explaining the causes of different emotions (Losh and Capps, 2006). Explanations of emotions reside at the heart of a good narrative, providing meaning to experience and connecting the narrator to conversational interlocutors (Bamberg and Damrad-Frye, 1991; Bruner, 2004; Labov and Waletzky, 1967). Lacking such essential elements, narratives from individuals with ASD may fail to achieve key social-communicative functions. Therefore, research characterizing narratives in less-structured contexts is crucial to better understanding the manifestation of these deficits in daily interactions.

This study importantly expands on prior work by examining narratives elicited by emotionally salient images from the TAT. Beaumont and Newcombe (2006) previously found that adults with ASD employed significantly fewer causal explanations of mental states in response to selected scenes from the TAT. In this study, we additionally assessed story structure, use of complex syntactic devices, idiosyncratic elements and expressions reflecting anxiety or task difficulty. By incorporating more comprehensive hand-coding methods to evaluate TAT narratives, we aimed to (1) characterize how difficulties with these aspects of narration manifest interpersonally and (2) establish a comprehensive basis for evaluating the effectiveness of the computational method, LSA, in capturing key indices of narrative ability in ASD.

**Table 1.** Group descriptive characteristics.

| Group | Age (SD)<br>Range | Sex (M:F) | FSIQ (SD)<br>Range | VIQ (SD)<br>Range | PIQ (SD)<br>Range |
|---|---|---|---|---|---|
| ASD ($n = 19$) | 24.22 (9.48)*<br>15.9–57.5 | 14:5* | 105.53 (18.74)*<br>65–131 | 104.37 (18.27)**<br>58–132 | 105.36 (19.64)<br>59–131 |
| Control ($n = 14$) | 19.11 (2.20)<br>15.7–25.6 | 5:9 | 119.50 (10.93)<br>91–135 | 122.07 (10.34)<br>98–138 | 112.57 (11.24)<br>87–127 |

SD: standard deviation; FSIQ: Full Scale IQ; VIQ: Verbal IQ; PIQ: Performance IQ; ASD: autism spectrum disorder.
*$p < 0.05$; **$p < 0.01$.

## LSA

LSAs have been applied extensively in psychological research, including modeling of word learning (Landauer and Dumais, 1997), describing couples' discourse (Babcock et al., 2014) and extraction of clinical information from psychiatric narratives (Cohen et al., 2008). LSA identifies the semantic similarities of words based on their statistical co-occurrence with other words in a "semantic space," typically a large corpus of diverse readings (e.g. general reading through college level). LSA can extend similarity at the word level to calculate semantic similarities of bodies of text, providing a similarity score ranging from near –1 (least similar) to 1 (most similar). Therefore, LSA shows potential as an objective, quantitative measure of narrative ability that may be applied to both large-scale research studies of narrative and clinical settings, where detailed hand coding is not feasible. Losh and Gordon (2014) applied LSA to narratives from individuals with ASD, showing that LSA effectively discriminates narrative quality in relatively structured contexts. This study expands on this work by applying LSA to stories told based on scenes from the TAT and examining associations between LSA and hand-coding variables.

## Methods

### *Participants*

Participants included 19 individuals with ASD and 14 TD controls who were participating in a larger family study that included families affected by ASD and control families without a history of ASD. Participants were eligible to enroll in the study if English was the first and primary language. Individuals with ASD were included if they had a Full Scale IQ (FSIQ) greater than 60, and individuals with TD were excluded if there was evidence of cognitive impairment (i.e. IQ less than 80). Additionally, individuals with ASD were required to have a previous clinical diagnosis, confirmed by administration of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2001) and/or the Autism Diagnostic Interview–Revised (ADI-R; Lord et al., 1994). Exclusion criteria for controls included a personal or family history of ASD or related disorders (e.g. fragile X syndrome, dyslexia). Individuals with ASD were recruited through advocacy groups, schools, health clinics and participant registries. Controls were recruited from the local University community. Several efforts were made to increase the diversity of the control sample, including advertising with community organizations serving families, schools and at restaurants and stores in the local University community and surrounding city, but many participants were affiliated with the University. All participants provided informed consent, and procedures were approved by Northwestern University Institutional Review Board (IRB #: STU00036069). Table 1 summarizes participant demographics. FSIQ, Verbal IQ (VIQ) and Performance IQ (PIQ) were derived from the Wechsler Abbreviated Scale of Intelligence (WASI) or the Wechsler Adult Intelligence Scale (WAIS)—Third or Fourth Editions (Wechsler, 1997, 1999, 2008). Given that participants were recruited as part of a larger family study focused primarily on relatives of individuals with ASD (where age and IQ of affected individuals were not primary recruitment criteria), and other noted recruitment strategies above, group differences in IQ, age and sex were present. However, as described in Analysis Plan and Results sections, these factors were examined in relationship to results and were not found to impact findings. Additionally, a primary focus of analyses concerned within-group associations between hand-coded and computationally derived narrative scores, rendering group differences of participant characteristics of somewhat less concern, particularly given their lack of impact on findings.

### *Procedure*

Following Paul et al. (2004), six TAT scenes were selected to reflect a range of emotional content; for example, one of the more "basic" images contained a boy looking at a violin with a clear sad expression, while a more complex scene included several characters overlooking a surgical table. Each scene was presented for 8 s on a computer monitor. Individuals were told that after viewing each image they were to tell a story with a beginning, middle and end, and to include what the characters were thinking, feeling and doing. If a participant's story did not contain all elements (thinking, feeling, doing and an ending), the examiner prompted for missing elements at the conclusion

of the participant's story. Prompts were included to assist participants in becoming familiar with the task, but to assess spontaneous narration utterances after prompts were not analyzed.

## Transcription

Narratives were recorded and then transcribed using ELAN, version 3.3 (Max Planck Institute for Psycholinguistics, 2002). Transcribers were trained to 80% agreement on word and utterance segmentation, and 20% of transcripts from each participant group were randomly selected to assess word and utterance reliability. Utterances were defined as groups of words that could not be divided without loss of meaning (Miller and Chapman, 2008). Mean word agreement was 96%, ranging from 85% to 100%. Utterance segmentation reliability was more variable (mean = 79%, range = 50%–100%) because narratives often contained few utterances, meaning that a single discrepancy substantially impacted reliability. Those transcripts with less than 75% utterance segmentation agreement were transcribed via consensus prior to analyses.

## Hand coding

Participants produced one narrative for each of the six selected TAT scenes. The narrative coding system employed in this study was derived from previous narrative research (e.g. Losh and Capps, 2003; Reilly et al., 1990, 1998). In total, 20% of participants in each group were randomly selected to assess coding reliability. The average inter-rater reliability (ICC) across groups was 0.68, signifying "good" agreement (Cicchetti, 1994). In addition, all transcripts were assessed by a third coder to check for coding inconsistencies or errors prior to analyses.

*Story length.* Story length was quantified as the number of words and utterances in each narrative. The total word count for each narrative was calculated using the Linguistic Inquiry and Word Count (LIWC) software (Pennebaker et al., 2001).

*Prompts.* The number of prompts a participant required to elicit descriptions of what characters were thinking, feeling or doing, and inclusion of an ending was tallied.

*Story structure.* Following Stein and Glenn (1979), story structure elements included major settings (e.g. the introduction of a character), minor settings (e.g. social, physical or temporal context), initiating events, a character's internal response, attempts to resolve, direct consequences and reaction to direct consequences. A "complete narrative episode" included an initiating event, an attempt to resolve *and* a direct consequence.

*Complex syntax.* The use of coordinate clauses, verb complements, relative clauses, passive clauses and adverbial clauses was totaled to assess the frequency and types of complex syntax devices employed (Reilly et al., 1998).

*Evaluative devices.* The use of simple affective states (e.g. happy, sad) or complex affective states (e.g. guilty, proud) and behaviors, cognitive states and behaviors, causal explanations of affect or cognition, and causal statements unrelated to affect or cognitions were totaled (Reilly et al., 1990).

*Idiosyncratic features of narrative.* Features that detracted from narrative comprehensibility and overall quality were noted, including illogical, redundant or vague statements, asides, and semantic-syntactic errors (Landa et al., 1992).

*Expressions of anxiety/difficulty.* Instances in which participants resisted task completion or expressed difficulty, uncertainty or negativity related to the task were noted. Short pauses (up to 5 s) and long pauses (6 s or more) during narration were also assessed.

## LSA

LSA is a computational linguistic tool originally developed to model the semantic relationships between words for search engine technology. LSA determines the meaning of a word by identifying patterns of its occurrence with other words and then represents that word as a point in a high-dimensional vector space (typically having 300 dimensions). Thus, words that have similar meanings would be plotted close together within this space, whereas words with dissimilar meanings would be plotted further apart. Although the dimensions that make up this vector space do not map on directly to clear semantic features commonly employed in hand-coding schemes (e.g. "emotion terms"), studies have shown that these distances are correlated with human judgments of word similarity (Landauer and Dumais, 1997; Landauer et al., 2007). This technology can be extended to bodies of text, or in the case of this study, individual narratives.

In order to identify the statistical co-occurrence of words, LSA must first be "trained" on a large corpus of written text. For this study, similarity metrics were computed using the S-Space implementation of LSA (Jurgens and Stevens, 2010). The semantic space was developed from a large corpus of transcribed subtitles for movies and television shows, shown to better approximate word meaning as it occurs in more naturalistic, conversational contexts (Brysbaert and New, 2009). The meaning of a text is represented by the sum of the vectors of the words that it contains and therefore is also a point in the high-dimensional vector space.

For the purposes of this work, for each scene, a "gold standard" narrative was identified from all participants

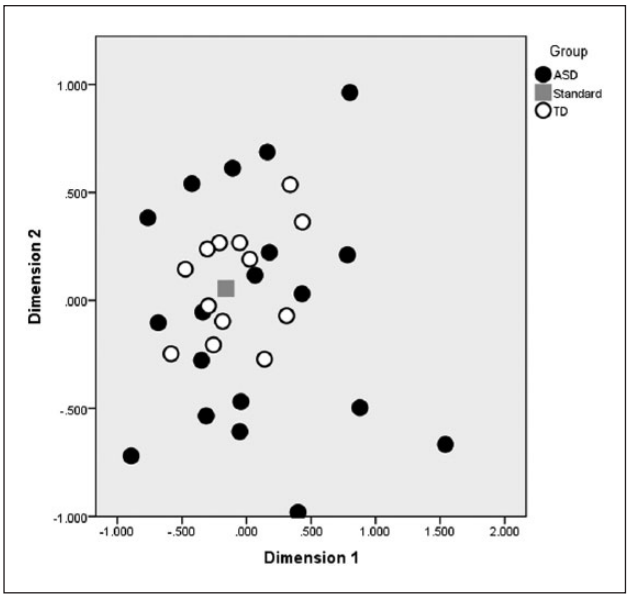


**Figure 1.** Distribution of LSA similarity scores for scene 5. This figure demonstrates the results of multi-dimensional scaling of LSA narratives to one another and in relationship to the gold standard, represented at the center of the space by a gray square. Of note, individuals with ASD, represented by black circles, are more dispersed than controls, represented by white circles, which are clustered more closely around the standard.

that was most similar to all other participant narratives. This approach resulted in a gold standard from the TD group for each slide. Then, for each slide, the similarity of each participant's narrative was compared to this standard text, represented by the distance in the vector space between the points representing the meaning of the two texts (more technically, the similarity is the cosine of the angle of the two vectors). This similarity was represented by a number ranging from −1 (not at all similar) to 1 (identical). Figure 1 provides a visual example of LSA applied to narratives from one scene of the TAT.

Of note, LSA does not model the *order of words*, and, given that it relies on the co-occurrence of words, high-frequency words (e.g. articles, prepositions) do not influence the ultimate similarity score (as these words co-occur with nearly every other word in the semantic space). Nevertheless, LSA provides an objective measure of the similarity of the semantic content of narrative, a metric that has been shown previously to distinguish individuals with ASD from TD controls (Losh and Gordon, 2014).

### *Analysis plan*

First, given group differences in sex and IQ, descriptive analyses were conducted to explore how these differences may have impacted results. Additionally, all analyses were replicated excluding two ASD participants with IQ less than 80 (more than 1 standard deviation below the mean), as described in the "Results" section.

To examine hand-coding differences, a series of planned comparisons were conducted to compare the average frequency of hand-coded elements across scenes, using analysis of covariance (ANCOVA), covarying for Verbal IQ (VIQ) or an inflated Poisson regression including group and VIQ as predictors, based on the distribution of data. ANCOVAs, also covarying for VIQ, compared the average LSA similarity metric of participants across scenes, and on each individual scene. In light of the small sample size, all ANCOVAs were also followed by non-parametric tests (Mann–Whitney *U*). To assess relationships between LSA and hand-coded elements, partial Pearson correlations were conducted, covarying for VIQ. For all analyses related to LSA, the gold standard for each slide was excluded. Although groups also differed in chronological age, chronological age was not included as a covariate given that prior research suggests that the narrative skills examined in this study are well developed by the adolescent period studied here (e.g. Reese et al., 2011) and the small magnitude of this difference (mean difference = 5.11 years). Eta-squared and adjusted mean differences were used to interpret effect sizes. Given small sample sizes and in order to avoid Type 2 error, we did not correct for multiple comparisons (which controls for Type 1 error); therefore, results should be considered preliminary.

## Results

### *Demographic variables and narration*

Initial analyses were conducted to explore whether differences in sex and IQ may have impacted results. For example, analyses of summary hand-coded variables were replicated excluding two ASD participants with IQ less than 80 (more than 1 standard deviation below the mean) and findings were consistent. Therefore, these participants were retained in order to maximize sample size and our ability to examine within-subject patterns between hand-coding and computational measures. Performance was also examined descriptively by sex, age and IQ. In the TD group, sex differences were negligible, ranging from a mean of 0.04 for an LSA outcome measure on a scene to 1.26 for idiosyncratic statements during narration. Therefore, it is unlikely that the inclusion of more females in the TD group drove significant findings. Within the ASD group, differences ranged from 0.02 to 2.64 (for an LSA outcome score on an individual scene and complex syntax, respectively). For both groups, the direction of difference was inconsistent. Together, these analyses suggest that while the following results should be interpreted with caution, differences in IQ and sex distribution did not impact reported findings.

### *Hand coding*

Table 2 presents mean differences, adjusted for VIQ, for the primary hand-coded categories. When interpreting

**Table 2.** Overall narrative performance on hand-coding measures.

| Hand-coding category | ASD *M* (SD) | TD *M* (SD) |
|---|---|---|
| Story clauses | 4.32 (2.58) | 5.61 (2.18) |
| Idiosyncratic aspects of narrative | 1.94 (1.88) | 2.32 (2.02) |
| Evaluative devices | 2.88 (2.48)* | 5.83 (2.60) |
| Complex syntax | 3.66 (3.13)* | 7.35 (3.40) |
| Expressions of anxiety and difficulty | 1.44 (1.44)* | 0.64 (0.60) |

TD: typically developing; ASD: autism spectrum disorder; *M*: mean; SD: standard deviation.
*$p < 0.05$.

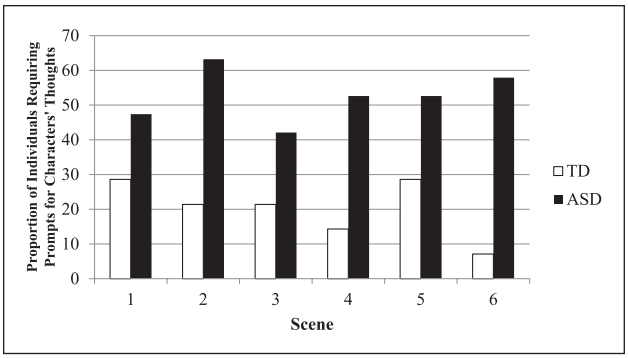


**Figure 2.** Proportion of individuals with ASD requiring prompts for characters' thoughts remained consistent across scenes.

eta-squared, 0.01 is considered a small effect, 0.06 is considered a medium effect and 0.14 is considered a large effect (Field, 2009). Overall, the narratives of individuals with ASD were similar to controls in many ways. Individuals with ASD did not differ from controls in their identification of basic story elements or idiosyncratic qualities across images. Nor did they differ from their TD peers in the average number of utterances across slides ($t(31) = 1.96$, $p = 0.059$). However, on average, the ASD group used fewer words than controls ($t(31) = 2.43$, $p = 0.021$) and included significantly fewer instances of complex syntax in their narratives ($F(1, 30) = 4.43$, $p = 0.04$, $\eta^2 = 0.10$; $U = 48.00$, $Z = -3.10$, $p = 0.001$; adjusted mean difference = 2.74), with particularly sparse use of coordinate and adverbial clauses ($F(1, 30) = 4.51$, $p = 0.04$, $\eta^2 = 0.11$; $U = 41.5$, $Z = -3.35$, $p < 0.001$, adjusted mean difference = 1.38; $F(1, 30) = 5.90$, $p = 0.02$, $\eta^2 = 0.15$; $U = 63.00$, $Z = -2.55$, $p = 0.01$, adjusted mean difference = 1.39). The ASD group also produced significantly fewer evaluative devices ($F(1, 30) = 4.80$, $p = 0.04$, $\eta^2 = 0.11$, $U = 49.00$, $Z = -3.06$, $p = 0.002$, adjusted mean difference = 2.22). Differences in evaluation were driven by the ASD group's limited use of complex affective states ($F(1, 30) = 3.71$, $p = 0.064$, $\eta^2 = 0.09$; $U = 52.00$, $Z = -2.96$, $p = 0.002$, adjusted mean difference = 0.70), affective behaviors ($F(1, 30) = 5.21$, $p = 0.03$, $\eta^2 = 0.14$; $U = 76.00$, $Z = -2.34$, $p = 0.04$, adjusted mean difference = 0.16) and causal explanations of cognition ($F(1, 30) = 3.08$, $p = 0.09$, $\eta^2 = 0.09$; $U = 75.50$, $Z = -2.33$, $p = 0.04$, adjusted mean difference = 0.18).

Additionally, the ASD group required significantly more prompts throughout the task (Wald Chi-Square = 12.78, $p < 0.01$), showing particular reliance on prompts to identify what characters were thinking (Wald Chi-Square = 6.40, $p = 0.01$) and how the story ended (Wald Chi-Square = 5.39, $p = 0.02$). As demonstrated in Figure 2, these differences persisted across scenes despite increased familiarity with the task. In line with these findings, narratives from individuals with ASD were also characterized by more frequent expressions of anxiety and the perceived difficulty of the task ($F(1, 30) = 8.04$, $p = 0.008$, $\eta^2 = 0.21$; $U = 96.0$, $Z = -1.36$, $p = 0.19$; adjusted mean difference = 1.3), including short and long pauses ($F(1, 30) = 8.79$, $p = 0.01$, $\eta^2 = 0.22$; $U = 89$, $Z = -1.64$, $p = 0.11$; adjusted mean difference = 1.1; $F(1, 30) = 12.76$, $p = 0.001$, $\eta^2 = 0.30$; $U = 70$, $Z = -2.75$, $p = 0.02$; adjusted mean difference = 0.2) and explicit expressions of task difficulty ($F(1, 30) = 4.43$, $p = 0.04$, $\eta^2 = 0.13$; $U = 106.5$, $Z = -1.44$, $p = 0.34$; adjusted mean difference = 0.1), although these comparisons were not confirmed by non-parametric tests. In the ASD group, more frequent prompting throughout the task was negatively correlated with evaluation ($r = -0.82$, $p < 0.001$), complex syntax ($r = -0.73$, $p < 0.001$), story clauses ($r = -0.81$, $p < 0.001$) and idiosyncratic aspects of narration ($r = -0.67$, $p < 0.001$); for controls, a greater number of prompts were negatively correlated with evaluation ($r = -0.71$, $p < 0.01$) and story clauses ($r = -0.58$, $p < 0.05$).

## LSA

Narratives from the ASD group were significantly lower in semantic similarity than controls on average ($F(1, 30) = 7.98$, $p = 0.008$, $\eta^2 = 0.12$; $U = 28.00$, $Z = -3.83$, $p < 0.0001$; adjusted mean difference = −0.104), and, in particular, in response to scene 1 and scene 5 ($F(1, 29) = 6.68$, $p = 0.015$, $\eta^2 = 0.15$; $U = 37$, $Z = -3.32$, $p = 0.001$; $F(1, 29) = 5.24$, $p = 0.03$, $\eta^2 = 0.11$; $U = 40$, $Z = -2.59$, $p = 0.010$) as illustrated in Figure 3.

Table 3 includes sample narratives designated by LSA with high and low similarity scores relative to standard narratives. In the ASD group, semantic similarity was positively correlated with key indices of narrative quality: total story clauses ($r = 0.62$, $p = 0.006$), complex syntax ($r = 0.60$, $p = 0.009$) and evaluative devices ($r = 0.59$, $p = 0.01$). Semantic similarity was also negatively correlated with total prompts ($r = -0.62$ $p = 0.005$), and driven by a negative relationship with total prompts for thinking ($r = -0.58$, $p = 0.01$).

## Discussion

Narration plays a pivotal role in social interaction. Prior findings suggest that narrative difficulties comprise a core component of social language deficits in ASD, in that difficulties formulating emotionally salient narratives, particularly in open-ended naturalistic contexts, limit participation in social-communicative interactions. This study examined narrative abilities of individuals with ASD using an open-ended, emotionally evocative narrative elicitation task that better approximates typical social interactions, relative to structured storybook tasks used in prior research. Additionally, we combined both hand-coding and computational linguistic methods, allowing for comprehensive assessment of narration and further validation of a promising computational method for assessing complex language abilities in ASD.

Overall, findings replicated prior work, indicating that the narratives of individuals with ASD are characterized by reduced use of complex syntax, evaluation and increased explicit expressions of task difficulty and need for structure. Consistent with detailed hand-coding methods, LSA detected differences in narratives between individuals with ASD and controls and was highly correlated with several hand-coding measures. Together, these results build substantially on prior research by identifying key narrative differences that may manifest in everyday interactions and further establishing the effectiveness of computational tools in capturing such differences, with both clinical and empirical applications.

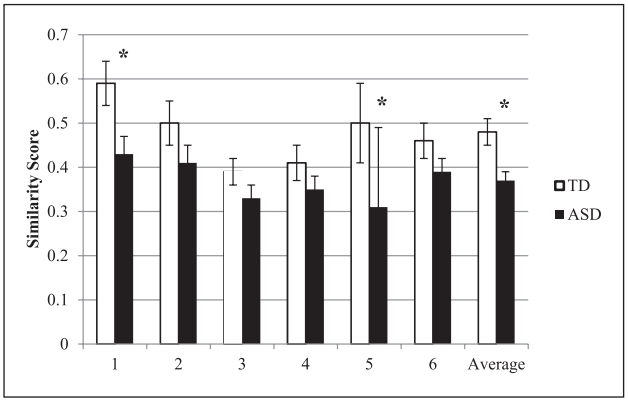


**Figure 3.** Group differences in average LSA similarity scores. *$p < 0.05$. Greater scores indicate greater similarity to a gold standard narrative.

Consistent with prior literature (Beaumont and Newcombe, 2006; Capps et al., 2000; Hogan-Brown et al., 2013; Tager-Flusberg and Sullivan, 1995), hand-coding results suggest that even in less-structured contexts, individuals with ASD are capable of producing narratives that include basic elements such as a beginning, middle and end; main characters; and identification of basic thoughts and emotions. However, important differences were also detected. Whereas the number of utterances was comparable across groups, narratives from the ASD group contained fewer words and less complex syntax; of note, individuals with ASD incorporated an average of nearly three fewer complex syntactic devices relative to controls, even after controlling for verbal abilities. Complex syntax is an important tool for integrating events into temporal-causal frameworks that provide critical structure and coherence to a narrative (Berman and Slobin, 1994). For example, in a scene depicting a woman sitting on a couch with a man looking over her shoulder, a control participant used verb complement, coordinate, and adverbial clauses to bind events in time and to establish causal relationships: "*When* Michael came home he thought it'd be funny *to scare* his wife *so* he walked in quietly into the house *and* saw his wife reading a book on the couch." Describing this same scene, an individual with ASD used only coordination, simply stating, "So a woman was resting on the couch *and* then a guy came up to her and threatened her." This lack of complex syntax resulted in a comparatively impoverished description of events, with limited causal and temporal integration of story elements.

**Table 3.** Examples of high and low semantic similarity narratives.

| Gold standard | High semantic similarity | Low semantic similarity |
|---|---|---|
| *TD Participant* (19 years old, VIQ = 122) | *ASD Participant* (18 years old, VIQ = 97, similarity = 0.50) | *ASD Participant* (15 years old, VIQ = 106, similarity = 0.21) |
| Uh so this story is about a little boy who intends to become a doctor and his dad is a surgeon so he's performing surgery on or this body. And the boy is very interested. And so he's observing because he wants to be just like his father. And the father is feeling very um anxious because he has to perform the surgery and very um concentrated on his work and the story ends with the boy becoming a famous doctor. | So I should I start now? so um a younger boy um um maybe ten maybe. His dad's a doctor. And his dad badly no his dad's a surgeon. his dad badly wants him to become a surgeon. But thoughts like you know about body parts and about blood just uh just disgusts him so much. Yeah so that's what he's thinking about. | Like basically um the doctor was cutting open someone and then like that person's brother or dad or someone was turned away, and yeah. |

TD: typically developing; ASD: autism spectrum disorder; VIQ: Verbal IQ.

Replicating prior work, the ASD group also failed to incorporate evaluation in their narration (Beaumont and Newcombe, 2006; Capps et al., 2000; Losh and Capps, 2003; Tager-Flusberg and Sullivan, 1995). Evaluative devices, such as explanations of characters' thoughts, are critical to infusing narration with a psychological perspective (e.g. "he decided that he wanted to be a surgeon *because* … at one of his factory jobs he had witnessed …"; Bamberg and Damrad-Frye, 1991; Berman and Slobin, 1994). Whereas individuals with ASD did not differ in their ability to identify and explain basic affective states, they integrated fewer complex emotions and provided fewer explanations of characters' thoughts, as has been demonstrated in previous studies (Capps et al., 2000; Losh and Capps, 2003; Tager-Flusberg and Sullivan, 1995). Therefore, these results confirm difficulty with evaluation as a core aspect of narrative impairment in ASD.

Individuals with ASD also required a greater level of scaffolding to produce narratives, likely restricting their capacity to narrate in everyday interactions where such extensive support from an interlocutor is unlikely. Strikingly, by the concluding image 58% of individuals with ASD required a prompt for what characters were thinking, compared to only 7% of controls. Notably, individuals with ASD paused frequently during their narrations and expressed clear discomfort with the task (e.g. "I don't like thinking of … what somebody's thinking"). Anxiety associated with narrative formulation likely contributes to the widely observed difficulties with narration experienced by individuals with ASD in social settings and is consistent with some prior research (Losh and Capps, 2003). Therefore, increasing comfort and confidence with narration is likely to be an important goal of language interventions for individuals with ASD.

In contrast to prior findings (Losh and Capps, 2003; Loveland et al., 1990; Norbury and Bishop, 2003), individuals with ASD did not differ in idiosyncratic aspects of narrative, such as off-topic comments or semantic errors. However, review of transcripts revealed qualitative differences in idiosyncratic elements characteristic of each group. The most common error among control participants was inconsistent use of tense, whereas the most common errors committed by individuals with ASD included incorrect word use, such as "boringly," or odd phrasing, such as "contemplating about." Therefore, while the groups did not differ in overall idiosyncratic qualities, such features may interfere with coherent narration among individuals with ASD.

Together, hand-coding results offer a rich picture of narrative differences in ASD that both confirm and extend prior findings, with implications for targeted interventions. Equally important were findings that LSA successfully differentiated groups, and correlated highly with key hand-coded measures of narrative competence, as previously demonstrated in a younger group of individuals with ASD using different narrative contexts (Losh and Gordon, 2014). Of note, LSA differences do not simply reflect differences in the length of narratives, as individuals with ASD did not differ from controls in number of utterances. Rather, LSA was associated with the use of sophisticated narrative devices such as complex syntax and evaluative language, suggesting that this tool may be useful in capturing even more subtle impairments expressed among very high-functioning individuals with ASD. Thus, LSA shows potential as a quantitative measure of global narrative quality that may be fruitfully applied to further research characterizing communicative impairments in ASD, as well as in clinical practice.

Currently, tools for the objective, valid and reliable assessment of social language use in ASD in clinical contexts are limited, particularly for higher functioning individuals who do not display obvious differences in structural language abilities. Standardized tests can often miss important differences that are evident in naturalistic settings, and informant reports, while important for capturing abilities in naturalistic contexts, may suffer from subjectivity. Given the centrality of narration to everyday interaction and impairments observed in naturalistic contexts despite intact language in high-functioning individuals in ASD, accurate assessment of narrative ability is critical for intervention planning. Clinical evaluations or hand coding of narratives are valid assessment methods that offer a comprehensive assessment of narrative features, but are difficult to accomplish quickly and reliably.

LSA offers considerable advantages over existing assessment tools in clinical settings in that it provides a rapid, objective and empirically derived quantitative assessment of complex language skills. LSA may be particularly useful in clinical contexts, not only for quantifying impairment in this essential communicative skill relative to a gold standard but also measuring variation in response to intervention. Findings that LSA was positively related to key hand-coded measures of narrative ability, replicating results from a prior investigation examining different narrative contexts in younger children with ASD (Losh and Gordon, 2014), suggest that LSA shows promise as a global tool to detect challenges in narration that are common in ASD. Such a tool could also prove important for neurobiological and genetic studies, where quantitative measures of complex traits that can be assessed across different ages and ability levels can increase power to detect gene–brain–behavior relationships (Gottesman and Gould, 2003; Greenwood et al., 2013). Application of LSA to language samples in such studies would provide a quantitative measure that could be applied across a large group of individuals with ASD and unaffected relatives to capture subtle variation in language that may map onto underlying genetic mechanisms in a manner far more feasible than applying hand coding.

It will be important to replicate these findings with larger samples and across additional types of narrative stimuli that may help to further define the profile of narrative strengths and weaknesses in ASD. Additionally, future work should aim to further refine the use of LSA in ASD, which holds strong potential as a tool for capturing complex language ability in large-scale empirical research and in clinical contexts, where hand coding is not feasible because of the extensive resources and expertise required. In particular, it will be important to examine sex differences in narration in ASD, as the distribution of the current sample did not allow for group comparisons by sex. Despite the many benefits of LSA, there are some important limitations to consider concerning its applications. For instance, LSA does not account for word order, which can clearly impact comprehensibility in conversation, although hand coding suggests that this was not an area of difficulty in the ASD group. The applications of LSA in response to diverse language contexts, and conversation in particular, should be explored. Examining the sensitivity of LSA in measuring changes in narrative ability will also be important to evaluate whether this tool might be useful in charting developmental growth or measuring response to intervention. In conclusion, results underscore key narrative impairments that characterize ASD and support the promise of LSA as a valuable tool for the objective, quantitative assessment of complex language abilities in research and clinical contexts.

## Acknowledgements

The authors thank Jan Misenheimer, John Sideris and all project staff for their contributions and are particularly grateful to the participants in this study.

## Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by grants from the National Institute on Deafness and Other Communication Disorders (1R01DC010191-01A1) and the National Institute of Mental Health (1R01MH091131-01A1; 3R01MH091131-03S1).

## References

Babcock M, Ta VP and Ickes W (2014) Latent semantic similarity and language style matching in initial dyadic interactions. *Journal of Language and Social Psychology* 33(1): 78–88.

Bamberg M and Damrad-Frye R (1991) On the ability to provide evaluative comments: further exploration of children's narrative competencies. *Journal of Child Language* 18: 689–710.

Beaumont R and Newcombe P (2006) Theory of mind and central coherence in adults with high functioning autism or Asperger syndrome. *Autism* 10(4): 365–382.

Berman R and Slobin D (eds) (1994) *Relating Events in Narrative: A Cross-linguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum.

Bruner JS (2004) Life as narrative. *Social Research* 71(3): 691–710.

Brysbaert M and New B (2009) Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41: 977–990.

Capps L, Losh M and Thurber C (2000) The frog ate a bug and made his mouth sad: narrative competence in children with autism. *Journal of Abnormal Child Psychology* 28: 193–204.

Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 6(4): 284–290.

Cohen T, Blatter B and Patel V (2008) Simulating expert clinical comprehension: adapting latent semantic analysis to accurately extract clinical concepts from psychiatric narrative. *Journal of Biomedical Informatics* 41(6): 1070–1087.

Colle L, Baron-Cohen S, Wheelwright S, et al. (2008) Narrative discourse in adults with high functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders* 38: 28–40.

Elvevag B, Foltz PW, Weinberger DR, et al. (2007) Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. *Schizophrenia Research* 93(1–3): 304–316.

Field A (2009) *Discovering Statistics Using SPSS: Third Edition*. London, England: SAGE Publishing.

Gottesman I and Gould T (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. *American Journal of Psychiatry* 160: 636–645.

Greenwood TA, Swerdlow NR, Gur RE, et al. (2013) Genome-wide linkage analyses of 12 endophenotypes for schizophrenia from the consortium on the genetics of schizophrenia. *American Journal of Psychiatry* 170(5): 521–532.

Hogan-Brown AL, Losh M, Martin GE, et al. (2013) An investigation of narrative ability in boys with autism and fragile X syndrome. *American Journal on Intellectual and Developmental Disabilities* 118(2): 77–94.

Jurgens D and Stevens K (2010) The S-Space package: an open source package for word space models. In: *Proceedings of the ACL 2010 system demonstrations*, Uppsala, 13 July.

Labov W and Waletzky J (1967) Narrative analysis. In: Labov W, Cohen P, Robins C, et al. (eds) *A Study of the Non-standard English of Negro and Puerto Rican Speakers in New York City*. New York: Columbia University Press, pp. 286–338.

Landa R, Piven J, Wzorek MM, et al. (1992) Social language use in parents of autistic individuals. *Psychological Medicine* 22(1): 245–254.

Landauer TK and Dumais ST (1997) A solution to Plato's problem: the latent semantic analysis theory of acquisition,

induction, and representation of knowledge. *Psychological Review* 104(2): 211–240.

Landauer TK, McNamara DS, Dennis S, et al. (2007) *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Lawrence Erlbaum.

Lord C, Rutter M and Le Couteur A (1994) Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders* 24(5): 659–685.

Lord C, Rutter M, DeLavore PC, et al. (2001) *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.

Losh M and Capps L (2003) Narrative ability in high-functioning children with autism or Asperger's syndrome. *Journal of Autism and Developmental Disorders* 33(3): 239–251.

Losh M and Capps L (2006) Understanding of emotional experience in autism: insights from the personal accounts of high-functioning children with autism. *Developmental Psychology* 42(5): 809–818.

Losh M and Gordon PC (2014) Quantifying narrative ability in autism spectrum disorder: a computational linguistic analysis of narrative coherence. *Journal of Autism and Developmental Disorders* 44(12): 3016–3025.

Loveland KA, McEvoy RE and Tunali B (1990) Narrative story telling in autism and Down's syndrome. *British Journal of Developmental Psychology* 8: 9–23.

Loveland KA and Tunali B (1993) Narrative language in autism and the theory of mind hypothesis: a wider perspective. In: Baron-Cohen S, Tager-Flusberg H and Cohen DJ (eds) *Understanding Other Minds: Perspectives From Autism*. Oxford, UK: Oxford University Press.

Max Planck Institute for Psycholinguistics (MPIF) (2002) *ELAN*. Nijmegen, The Netherlands. Available at: http://tla.mpi.nl/tools/tla-tools/elan/

Miller JF and Chapman RS (2008) *Systematic Analysis of Language Transcripts (SALT)*. Madison: University of Wisconsin-Madison.

Murray H (1943) *Manual of Thematic Apperception Test*. Boston, MA: Harvard University Press.

Norbury CF and Bishop DV (2003) Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders* 38(3): 287–313.

Ochs E and Capps L (2001) *Living Narrative: Creating Lives in Everyday Storytelling*. Cambridge, MA: Harvard University Press.

Paul LK, Schieffer B and Brown WS (2004) Social processing deficits in agenesis of the corpus callosum: narratives from the thematic apperception test. *Archives of Clinical Neuropsychology* 19: 215–225.

Pennebaker JW, Francis ME and Booth RJ (2001) *Linguistic Inquiry and Word Count* (LIWC). Mahwah, NJ: Lawrence Erlbaum.

Reese E, Haden CA, Baker-Ward L, et al. (2011) Coherence of personal narratives across the lifespan: a multidimensional model and coding method. *Journal of Cognitive Development* 12(4): 424–462.

Reilly J, Bates E and Marchman V (1998) Narrative discourse in children with early focal brain injury. *Brain and Language* 61: 335–375.

Reilly J, Klima E and Bellugi U (1990) Once more with feeling: affect and language in atypical populations. *Development and Psychopathology* 2: 367–391.

Stein NL and Glenn CG (1979) Directions in discourse processing. In: Freedle RO (ed.) *An Analysis of Story Comprehension in Elementary School Children*. Norwood, NJ: ABLEX Publishing, pp. 53–120.

Tager-Flusberg H and Sullivan K (1995) Attributing mental states to story characters: a comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics* 16(3): 241–256.

Wechsler D (1997) *Wechsler Adult Intelligence Scale-Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler D (1999) *Wechsler Abbreviated Scale of Intelligence* (WASI). London: Pearson Assessment.

Wechsler D (2008) *Wechsler Adult Intelligence Scale-Fourth Edition*. San Antonio, TX: The Psychological Corporation.