



Interpretations of meaningful and ambiguous hand gestures in autistic and non-autistic adults: A norming study

Brianna E. Cairney¹ · Stanley H. West¹ · Eileen Haebig² · Christopher R. Cox¹ · Heather D. Lucas¹

Accepted: 6 October 2023 / Published online: 27 November 2023
© The Psychonomic Society, Inc. 2023

Abstract

Gestures are ubiquitous in human communication, and a growing but inconsistent body of research suggests that people with autism spectrum disorder (ASD) may process co-speech gestures differently from neurotypical individuals. To facilitate research on this topic, we created a database of 162 gesture videos that have been normed for comprehensibility by both autistic and non-autistic raters. These videos portray an actor performing silent gestures that range from highly meaningful (e.g., iconic gestures) to ambiguous or meaningless. Each video was rated for meaningfulness and given a one-word descriptor by 40 autistic and 40 non-autistic adults, and analyses were conducted to assess the level of within- and across-group agreement. Across gestures, the meaningfulness ratings provided by raters with and without ASD correlated at $r > 0.90$, indicating a very high level of agreement. Overall, autistic raters produced a more diverse set of verbal labels for each gesture than did non-autistic raters. However, measures of within-gesture semantic similarity among the responses provided by each group did not differ, suggesting that increased variability within the ASD group may have occurred at the lexical rather than semantic level. This study is the first to compare gesture naming between autistic and non-autistic individuals, and the resulting dataset is the first gesture stimulus set for which both groups were equally represented in the norming process. This database also has broad applicability to other areas of research related to gesture processing and comprehension. The video database and accompanying norming data are available on the Open Science Framework.

Keywords Gesture · Communication · Language · Autism spectrum disorder · Video database

Introduction

Gestures are an important aspect of human communication. Humans begin gesturing as early as 10 months of age (Bates, 1979) and these early gestures serve as important precursors to later spoken language. A substantial amount of research has demonstrated that co-speech gesture enhances communication (Driskell & Radtke, 2003; Goldin-Meadow, 1999; Hostetter, 2011) and facilitates learning and memory (Beattie & Shovelton, 2001; Breckinridge Church et al., 2007; Cook et al., 2010; Thompson et al., 1998). Nonetheless, many questions remain about when and how these benefits occur (for

recent reviews, see Kandana Arachchige et al., 2021; Dargue et al., 2019). Among other examples, current evidence conflicts about the role of gesture meaningfulness in memory enhancement (So et al., 2012; Straube et al., 2014; Feyereisen, 2006; Kartalkanat, & Göksun, 2020; Levantinou & Navarretta, 2016), the circumstances under which incongruent gestures impair understanding of language (Wu & Coulson, 2007b; Habets et al., 2011), and the extent to which gestures impact early versus late stages of language processing (Kelly et al., 2004; Wu & Coulson, 2005; Wu & Coulson, 2007a).

Research on how gesture impacts cognition can inform understanding of autism spectrum disorder (ASD), a developmental disorder marked by deficits in social communication for which diagnostic criteria include difficulties with nonverbal communication, including but not limited to the understanding and use of gestures (American Psychiatric Association, 2013). For children and adults alike, the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012) directs clinicians to assess the quality and quantity of gesture use during social interaction. Compared to

✉ Brianna E. Cairney
bcairn2@lsu.edu

¹ Department of Psychology, Louisiana State University, 236 Audubon Hall, Baton Rouge, LA 70803, USA

² Department of Communication Sciences & Disorders, Louisiana State University, Baton Rouge, LA 70803, USA

their typically developing peers, autistic infants and toddlers show delays and decreases in gesture production—particularly deictic (i.e., pointing) gestures (Manwaring et al., 2018; Özçalışkan et al., 2016) and those with a communicative function (e.g., gestures directing attention to oneself or to an object to express a shared interest; Mishra et al., 2020; Watson et al., 2013). The frequency of gesture production is generally comparable between adolescents and adults with and without autism (de Marchena & Eigsti, 2010 and Lambrechts et al., 2014a) but there are nuanced differences related to the timing and motion of gestures. Adults with autism produce slower movements and more pauses between gestures compared to non-autistic adults, and such temporal features are associated with measures of motor cognition and social awareness (Trujillo et al., 2021). In some instances, autistic adults alter non-verbal behaviors to present as more neurotypical (i.e., masking; Cook et al., 2022) and tend to rely on gesture production to signal conversational turns to a greater extent than their non-autistic peers (de Marchena et al., 2019).

Research findings regarding gesture comprehension in autistic populations are inconsistent (Dimitrova et al., 2017; Lambrechts et al., 2014b; Clements & Chawarska, 2010; de Marchena & Eigsti, 2010; Silverman et al., 2017; Fourie et al., 2020; Hubbard et al., 2012). Some studies report impaired gesture comprehension in children and adults with ASD, as evidenced by less accurate matching of meaningful gestures to appropriate pictures, words, or objects (Cossu et al., 2012). Atypical comprehension and integration of gesture with speech may be particularly impacted when nonverbal cues are communicative (e.g., pointing versus grasping; Aldaqré et al., 2016) or pro-social (e.g., pointing to share an experience versus pointing to request a desired item; Clements & Chawarska, 2010). However, other studies using similar tasks have found no comprehension differences (Dimitrova et al., 2017; Adornetti et al., 2019).

Many factors, such as participant age, symptom type and severity, and task demands, may account for these and other discrepancies. However, heterogeneity in the stimuli used across studies likely contributes substantially to conflicting findings. Agostini et al. (2019) outlined several sources of stimulus-based variability that may impact the results of gesture comprehension studies, including the overall comprehensibility of the gestures, the types of gesture (e.g., pantomimes versus emblems), and whether the videos or images depict the actor's entire face and body, torso only, or just the hands. Other important considerations for studies pertaining to ASD may be the complexity of the movement involved in the gesture (Trujillo et al., 2021) and presence or absence of face visibility (Dawson et al., 2005), given evidence that people with ASD spend less time than people without ASD viewing the faces of actors when observing and attempting to imitate gestures (Vivanti et al., 2008).

We have developed a database of 162 well-characterized videos depicting hand gestures ranging in iconicity for use in research that have been normed for comprehensibility by

groups of participants with and without a self-reported diagnosis of ASD. During the norming process, adult participants from both diagnostic groups were asked to: (1) provide a numerical rating of the meaningfulness of each gesture, and (2) regardless of meaningfulness, provide a one-word label that best describes the gesture. The inclusion of ambiguous as well as meaningful gestures makes this stimulus set well suited for research on the role of semantics in gesture comprehension and/or imitation (e.g., “dual-route” theories of imitation; Stieglitz Ham et al., 2011; Bartolo et al., 2001). All gestures were edited to be of uniform duration. While the gesture stimuli are silent, the face and mouth are obscured from view to allow other researchers the option of embedding audio. The normed database of gestures is publicly available via the Open Science Framework repository (OSF), <https://osf.io/a3pg7/>.

In addition to producing a stimulus database for future research, the present study is the first to our knowledge to examine gesture naming and perceived meaningfulness of iconic and ambiguous gestures in people with and without ASD. Importantly, open-ended naming tasks may be sensitive to differences in gesture comprehension that are too subtle to impact performance on traditional matching tasks. Thus, along with the numerical meaningfulness ratings and set of verbal responses given by each group, we also provide overall- and group-specific measures for each gesture of: (1) the number of distinct responses generated during the naming task; (2) the information-theoretic measure of entropy, which is sensitive to the level of competition or dominance among possible responses; and (3) the mean semantic distance of each response from each other response based on corpus-derived semantic distance metrics. Together, these measures permit comparisons of response heterogeneity at both the lexical level (number of distinct responses generated) and the semantic level (the extent to which the responses generated are similar in meaning).

Finally, to further enhance the utility of this resource, we include: (1) measures that quantify the overall amount of movement involved in the gesture, and (2) the age-of-acquisition (based on existing norms; Kuperman et al., 2012) of the labels given by each participant to each gesture. These ancillary measures will benefit researchers wishing to control for motion variability and/or tailor their use of the stimulus set for child or adolescent participants.

Methods

Participants

Participants were recruited on Prolific (<http://www.prolific.co>), a recruitment platform for online research. Prolific's stand-alone demographic profiles and screening tools were used to selectively advertise the study to participants who identified as being between the ages of 18 and 40, living in the

United States, and native and primary speakers of English. An additional screener was used to advertise to a specific number of participants whose profile endorsed having received a formal clinical diagnosis of autism spectrum disorder as either a child or an adult (ASD group), as well as an equal number of participants who self-reported that they had never received an autism diagnosis (non-autistic or NA group). Importantly, the nature of Prolific's screening capabilities is such that participants are unable to know why any specific study has been made available to them, and autism was not mentioned as an inclusion criterion in our study description. These procedures are in line with empirically supported best practices for minimizing instances of participants misrepresenting themselves to gain access to the study or receive compensation, which is a concern for online research (Chandler & Paolacci, 2017). A detailed description of our recruitment and screening processes is provided in the Supplemental Methods.

Three versions of the task were available, each containing a different set of gestures. Participants were allowed to complete more than one version but were not allowed to complete the same version more than once. Overall, we received 182 submissions from 128 individuals who reported having a diagnosis of autism and 137 submissions from 129 individuals who reported not having a diagnosis. Submissions were excluded from analyses if they were incomplete, included low-effort responses, or if the participant provided demographic information that conflicted with their Prolific profile (see Supplemental Methods for details). This process resulted in the exclusion of 17 submissions from 15 participants in the NA group and 62 submissions from 45 participants in the ASD group (Tables S1 and S2), leaving a total of 120 submissions per group (40 for each task version) that came from 114 and 83 unique NA and ASD participants respectively. Table 1 depicts demographic information and group comparisons for age, gender, race, and level of education. The only significant difference in any of these metrics with respect to autism status was race, which was driven by a higher proportion of Asian participants in the NA group.

This study was approved by the Louisiana State University Institutional Review Board and participants provided informed consent prior to participating.

Stimuli

A set of 162 hand gestures were silently filmed on a 13-inch MacBook Pro using Photo Booth (Apple Inc). An actor, seated on a chair, was visible only from the neck to waist. Of the 162 gestures, 108 were categorized as iconic gestures and the remaining 54 as nonsense gestures. We use the term “iconic” to refer to gestures that tend to evoke specific, identifiable action concepts when viewed in isolation (e.g., in the absence of concurrent speech or other disambiguating context), and “nonsense” to refer to gestures that are perceived as relatively meaningless or ambiguous in isolation

(see Fig. 1 for examples). These classifications were made based on norming data from a pilot study that we conducted, which is described in detail in Supplemental Methods. It is important to note that, both during piloting and in the data presented here, a wide range of meaningfulness values are evident within each category. Thus, these labels should be viewed as relative rather than absolute, and we include them primarily as informative descriptors for researchers who wish to use this stimulus set. Meaningfulness ratings and free response labels are available on OSF for all gestures.

Videos were edited into 8.5-s clips using Adobe Premier Pro (Adobe Inc). As shown in Fig. 1, the timing of every video clip was standardized, such that the gesture was initiated 2.5 s into the clip, lasted 2.5 s, and ended with 3.5 s of stillness. To ensure that every gesture lasted precisely 2.5 s, some of the gestures were slightly sped up or slowed down from the original recorded speed. Each video clip was exported to .mp4 format with a 540-pixel resolution using H.264 (AVC) compression. The frame rate was 24 frames/second, or a total of 204 frames per video. Gestures were randomly divided into one of three groups for norming, each consisting of 36 iconic and between 17 and 19 nonsense gestures.

Procedure

After enrolling in the study on Prolific, participants were directed to Qualtrics (Qualtrics, Provo, UT) to complete the task. After reading task instructions, participants were presented with one gesture video at a time. For each video, participants were asked to respond on a 0–4 scale to the question, “*How meaningful did you find this gesture?*”, with anchor points at zero (not meaningful), 2 (somewhat meaningful), and 4 (completely meaningful). Next, participants were asked, “*If you had to choose one word to describe this gesture, what would it be?*” and were instructed to limit their response to one word. Four attention check trials were also shown. During these trials, the actor was present and seated in the same position without movement. A text overlay appeared two seconds into the video that read “ATTENTION CHECK” and instructed the participant to assign a meaningfulness rating of 4 to the associated trial.

Participants could watch each video more than once and responses were untimed. At the end of the study, participants were asked to complete a brief optional demographics form, which included questions about age, gender, education, native language, American Sign Language (ASL) fluency¹,

¹ We asked participants whether they knew ASL in case any of our gestures were unintentionally similar to signs. ASL signs are distinctly different from hand gestures in that they are structured similarly to spoken language and have modality-specific differences (Özyürek & Woll, 2019) but they can sometimes share physical or motoric similarity with iconic gestures. Although 21 autistic and 29 non-autistic participants reported some exposure to ASL, none were fluent and thus included in all analyses.

Table 1 Participant characteristics

| | Autistic | Non-autistic |
|------------------------|------------|--|
| Unique participants | 83 | 114 |
| Age | | $t(179.42) = 1.90, p = 0.06, d = 0.27$ |
| Range | 18 – 40 | 18 – 40 |
| <i>M</i> (<i>SD</i>) | 27.5 (6.1) | 25.8 (6.3) |
| Gender | | $\chi^2 = 2.60, p = 0.28, \phi = .1149$ |
| Female | 31 | 53 |
| Male | 47 | 58 |
| Non-binary | 5 | 3 |
| Race | | $\chi^2 = 15.46, p = 0.006, \phi = 0.28$ |
| White | 66 | 67 |
| Asian | 1 | 18 |
| Black | 4 | 7 |
| Mixed | 8 | 11 |
| Other | 2 | 7 |
| Data unavailable | 2 | 4 |
| Education | | $\chi^2 = 3.30, p = 0.67, \phi = 0.13$ |
| Highschool/GED | 14 | 16 |
| Some college | 42 | 55 |
| Four-year degree | 16 | 25 |
| Some graduate school | 2 | 4 |
| Graduate degree | 6 | 13 |
| None of the above | 3 | 1 |

Note. Gestures were divided into three sets for the rating study, and each set was rated by 40 autistic and 40 non-autistic participants (see Methods). The total number of unique participants per group is less than 120 because some participants rated more than one set. Age, gender, and education level data were acquired via a brief demographics questionnaire at the end of the experiment. Race data were obtained from Prolific (specific ethnicity data were not available). An independent samples *t* test was conducted compare age between the ASD and NA groups, and Pearson's chi-squared tests with simulated *p* values (based on 10,000 permutations) were conducted for gender, race, and education level. Groups did not differ on any measure except race. Post hoc comparisons revealed that this difference was driven by the higher proportion of Asian participants in the NA group. When comparing the groups on race after excluding the Asian subgroups, the groups did not statistically differ by race ($X(3, N = 172) = 3.26, p = 0.37$)

handedness, and ASD diagnosis. The mean time taken to complete the study was 30 minutes.

Free response lemmatization

Prior to statistical analysis, we lemmatized all 12,960 free responses using a combination of manual and automated procedures. Lemmatization is a process in which words are reduced to their base vocabular form. Verb tense, pluralization, and other inflections are removed (Jongejan & Dalanis, 2009). Lemmatizing words improves precision in text analysis because different variations of the same base word (e.g., breaking, broke, broken, breaks) can be interpreted as the same response (e.g., break).

Prior to lemmatization, misspelled responses were corrected using the spell-check function in Excel. All atypical responses were documented and reviewed by two experimenters. Homonyms were corrected when it was clear that the wrong word was provided (e.g., *peddle* was changed to

pedal if made in response to the gesture for pedaling). When participants provided more than one response, only the first word was saved (e.g., if the participant entered *drive/steer*, we saved *drive*). Words potentially perceived as offensive (e.g., curse words, sexual content ($n = 12$), non-words ($n = 3$), or responses that exceeded two words ($n = 16$) were excluded from analyses. We also excluded responses that appeared to be typos ($n = 1$), responses indicating technical problems ($n = 1$), “NA” entries ($n = 11$), and entries left blank ($n = 85$). All responses that indicated a lack of effort (e.g., “don’t know”, “unsure”, “meaningless”, etc.) were excluded ($n = 132$) unless they were given a meaningfulness rating of 3 or 4 or made in response to the gesture for *balancing*, *teetering*, *guessing*, or *shrugging* ($n = 76$). In the latter instance, responses such as “don’t know” or “unsure” were appropriate to the corresponding gesture and included in analyses. In total, these procedures resulted in the exclusion of 261 of the 12,960 responses, or 2.01% of the data. All spelling corrections and excluded responses are documented and available on our OSF page.

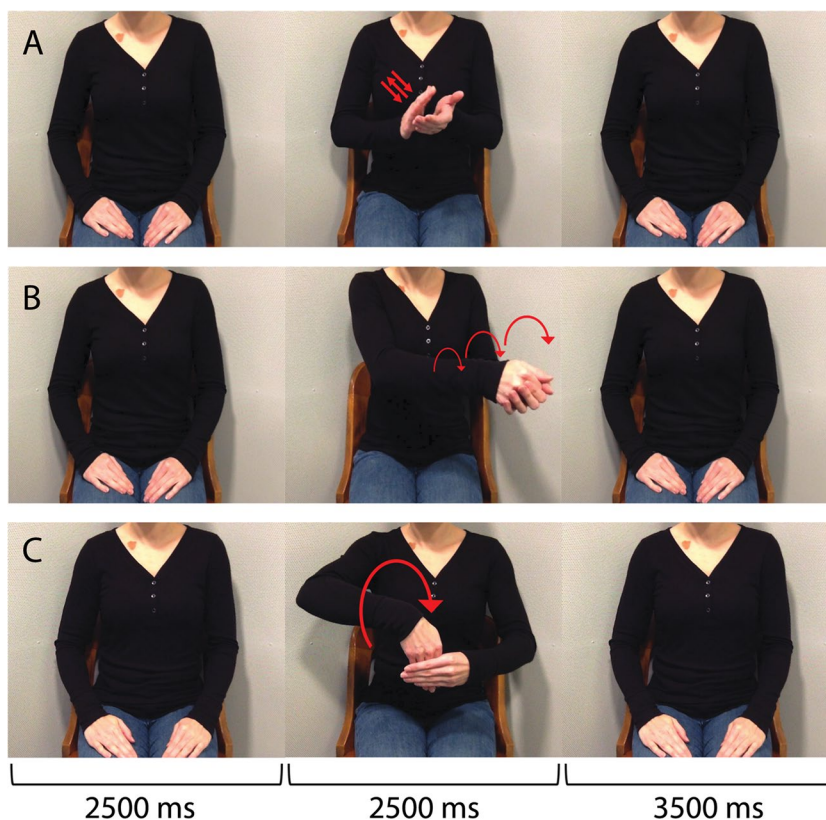


Fig. 1 Stills of example gesture video stimuli and stimulus timing. *Note.* **A** An iconic gesture intended to represent *clapping* or *applause*, **B** a nonsense gesture with no intended meaning, and **C** an iconic gesture intended to convey the word *insert* or *into*. Each gesture video

was edited into an 8.5-s clip. Gesture video clips began with the actor's hands resting in her lap for 2.5 s. Each gesture lasted precisely 2.5 s and the clip ended with 3.5 s of stillness.

We lemmatized the data based on the `hash_lemma` dictionary, which is a lemmatizing dictionary contained in the "lexicon" R package (Rinker, 2018a). This dictionary includes over 40,000 words and their corresponding lemma forms. We applied this dictionary to all free response data using the "textstem" R package (Rinker, 2018b), which is an automated text regularization program that changes words to their lemma form. The output of the program was then reviewed by two experimenters. One change in the lemmatized data was not appropriate (*feel* changed to *fee*) and corrected accordingly. We also accepted *cheers* as a salutation, rather than the lemma *cheer*, as this response matched the corresponding gesture.

Calculation of ancillary measures

Motion tracking Motion tracking was applied to each video using the software OpenPose (Cao et al., 2018). OpenPose is a motion tracking software that uses a deep neural net trained to identify human body poses in videos. We used OpenPose's "body25" model (which is recommended in the context of stimulus control; Trettenbrein & Zaccarella,

2021) to extract the *x*- and *y*- coordinates of keypoints on each wrist, elbow, and shoulder as well as on the top and bottom of the torso for each frame of each video (See Supplemental Fig. 1 for illustration). We then used the R package "OpenPoseR" (Trettenbrein & Zaccarella, 2021) to quantify the amount of bodily motion between adjacent frames. This process involves first computing the velocity of the individual keypoints along the *x* and *y* axis, and then computing the Euclidean norm of the sums of the velocity vectors (ENSVV), which yields a single value per frame representing the total change in motion relative to the prior frame.

Prior to calculating velocity, the motion tracking data were cleaned using OpenPoseR's `file_clean` function. This function identifies instances in which the model failed to fit a particular point or did so with low confidence (cutoff = 0.3) and imputes these values with the mean of the previous and consecutive frames. ENSVVs that were extreme outliers (values of > 8000) were also imputed. Finally, a third-order Kolmogorov–Zurbenko filter (span = 5) was applied to each gesture's sequence of ENSVVs to handle remaining high-frequency jitter due to sampling error. Filtering was implemented using the R package "kza" (Close et al., 2020).

Frame-by-frame ENSVVs for each video are available on OSF, as are the mean values each video for the pre-gesture time interval (second 0–2.5 or frames 2–60), the time interval containing the gesture (seconds 2.5–5 or frames 61–120) and post-gesture interval (seconds 5–8.5/frames 121–204). Overall, the mean ENSVV for iconic gestures was 89.61 per frame during the pre-gesture interval ($SD = 17.55$), 402.82 during the gesture interval ($SD = 110.87$), and 81.56 during the post-gesture interval ($SD = 12.66$). For nonsense gestures, the values were 91.29 (21.10), 476.72 (132.58), and 81.12 (17.04), respectively.

Response age-of-acquisition To estimate the age of acquisition (AoA) of the lemmatized versions of the words used to describe the gestures, we turned to a published set of norms for 30,121 English content words collected from people residing in the U.S. (Kuperman et al., 2012). It is important to note that the age at which a word that is used to describe a gesture is learned may differ somewhat from the age at which that word's meaning can be clearly conveyed by a specific gesture. For example, while the word “anger” has an AoA of six, older children and adults may be able to infer anger from a wider range of nonverbal cues than 6-year-olds. Nonetheless, given that ASD is a developmental disorder and often studied in children or adolescents, we include this measure as a starting point for researchers interested in identifying subsets of the gestures that may be appropriate for young research participants.

AoA ratings were available for 12,408 of the 12,960 (96%) responses in the dataset. There was a mean of 76.7 responses per gesture ($SD = 5.36$, range = 34–80). Per gesture, the number of responses provided by participants in the ASD group with AoAs available did not differ from the number provided by NA participants ($M_{ASD} = 38.24$, $SD_{ASD} = 2.71$; $M_{NA} = 38.36$, $SD_{NA} = 2.85$; $t(161) = 1.08$, $p = 0.28$, Cohen's $d = 0.04$), nor did the mean AoA of the responses given to the gestures by each group (means = 5.77 and 5.75 for the ASD and NA group, respectively; $SD = 0.86$ and 0.93 , $t(161) = 0.53$, $p = 0.60$, Cohen's $d = 0.02$). Overall, the mean AoA for responses given to iconic gestures was 5.70 ($SD = 0.93$, range = 3.55–7.98). For nonsense gestures, the mean AoA was 5.86 ($SD = 0.70$, range = 4.30–7.56).²

Analytic strategy

Statistical analyses served two main purposes. The first was to validate differences in perceived meaningfulness between the 108 gestures that were categorized *a priori* as “iconic”

² Information about the range of AoAs associated with responses to each specific gesture is available on the OSF page associated with this study.

and the 54 gestures categorized as “nonsense”. In addition to higher average meaningfulness ratings for the iconic relative to the nonsense gestures, we would also expect more consensus among raters in the free responses produced for iconic gestures, which should manifest as lower response diversity (e.g., fewer distinct words used to name each gesture), lower response entropy (more dominance of some responses over others), and higher semantic similarities among responses to a given gesture. Semantic similarity scores were calculated using Global Vectors for Word Representation (GloVe; Pennington et al., 2014), an unsupervised learning algorithm that extracts word vector representations from co-occurrence probabilities in natural language corpora. Procedures for calculating response diversity, entropy, and semantic similarity are described below.

Our second analysis goal was to compare the response profiles of participants with and without a diagnosis of ASD. Because group differences may vary at different levels of ambiguity, each variable of interest (meaningfulness, response diversity, response entropy, and response set semantic similarity) was analyzed using 2 (Gesture Category: Iconic vs. Nonsense) \times 2 (Group: ASD vs. NA) mixed factor ANOVAs. In addition, we computed across-gesture correlation coefficients between the values obtained for each variable from the ASD and NA samples separately for the iconic and nonsense gestures. This set of analyses examined the extent to which relative differences in the measures of interest among gestures were similar between the groups.

Response diversity

A response diversity score was calculated for each gesture within each group by dividing the number of unique responses provided by the total number of responses provided. This calculation only considered responses that appeared in the GloVe corpus.³

Response entropy

A score was calculated for each gesture within each group based on the information-theoretic measure of entropy (H ; Shannon, 1948). For a gesture that was assigned a total of i

³ Of the 12,699 one-word responses included in analyses, less than 1% ($n = 45$) did not appear in the GloVe corpus. There was no statistical difference between responses that were in the GloVe corpus versus all responses for entropy or diversity scores. To maintain consistency with semantic similarity scores (which can only be calculated from words in the corpus) we report entropy and diversity calculated from responses that appeared in the GloVe corpus. All norms are reported for each gesture in the OSF database, including the total number of responses provided and how many of those responses appeared in GloVe.

unique meanings, entropy is calculated using the following formula, in which R represents the number of unique meanings assigned to a given gesture and p_i is the proportion of participants who produced each unique meaning:

$$H = - \sum_{i=1}^R p_i \log_2(p_i)$$

This measure provides an index of how well the label elicited by a specific gesture can be predicted that takes into account both the number and distribution of responses. A higher entropy score indicates a relatively even distribution among the set of responses, whereas lower entropy scores occur when some responses are more dominant than others. As a hypothetical example, if the labels “draw”, “sketch”, “scribble”, and “doodle” were each produced 25% of the time in response to a given gesture, the response entropy for that gesture would be 2.0. By contrast, if the “draw” label was produced by 75% of raters, “sketch” by 15%, “scribble” by 8%, and “doodle” by 2%, the resulting entropy value would be 1.13, reflecting the dominance of some responses over others. Only responses present in the GloVe corpus contributed to entropy scores.

Response semantic similarity

A mean semantic similarity score was calculated for the set of unique responses given to each gesture within each group. Pairwise similarity values were calculated from a 300-dimension GloVe embedding. Using the “sim2” function of the “text2vec” R package (Selivanov et al., 2022), the cosine similarity was calculated for each pair of unique lemmatized responses provided for each gesture based on their corresponding vectors in GloVe. These pairwise values, which range from -1 to 1 , were averaged to yield a single measure of response semantic similarity for each gesture-group combination. Higher similarity values indicate that the responses provided by participants tended to have more closely related meanings, while lower similarity values identify gestures that elicited less related responses. For example, a gesture that received “yell”, “shout”, and “scream” as responses would receive a response similarity score of 0.61, whereas the response set “draw”, “shake”, and “tickle” would yield a score of 0.10.

Results

The means and standard deviations for each measure of interest (meaningfulness, response diversity, response entropy, and response semantic similarity) are described in Table 2 and Fig. 2, subdivided by group and gesture type. Across-gesture correlations are shown in Fig. 3.

Table 2 Means and standard errors of gesture norming metrics

| | ASD group | NA group |
|------------------------|-------------|-------------|
| Meaningfulness ratings | | |
| Iconic | 2.74 (0.06) | 2.75 (0.06) |
| Nonsense | 1.86 (0.08) | 1.77 (0.08) |
| Diversity | | |
| Iconic | 0.36 (0.02) | 0.32 (0.01) |
| Nonsense | 0.52 (0.02) | 0.50 (0.02) |
| Entropy | | |
| Iconic | 2.78 (0.01) | 2.56 (0.09) |
| Nonsense | 3.72 (0.09) | 3.64 (0.09) |
| Semantic similarity | | |
| Iconic | 0.17 (0.00) | 0.17 (0.01) |
| Nonsense | 0.15 (0.01) | 0.15 (0.01) |

Mean meaningfulness ratings, response diversity, response entropy, and response semantic similarity scores are subdivided by participant group and gesture type. ASD = autism spectrum disorder, NA = non-autistic. Standard errors in parentheses

Meaningfulness ratings

As expected, a significant main effect of gesture category on meaningfulness ratings emerged, $F(1, 160) = 91.27$, $p < 0.001$, $\eta_p^2 = 0.36$, such that gestures that were categorized *a priori* as iconic gestures were rated as more meaningful than those categorized as nonsense gestures (Fig. 2a). Meaningfulness ratings did not differ between groups, $F(1, 160) = 2.49$, $p = 0.12$, $\eta_p^2 = 0.02$. However, there was a significant condition \times group interaction, $F(1, 160) = 6.31$, $p = 0.01$, $\eta_p^2 = 0.04$. Follow-up t tests revealed that participants in the NA group found nonsense gestures less meaningful than those in the ASD group, $t(53) = 2.67$, $p = 0.01$, Cohen’s $d = 0.15$, whereas no group differences were present for iconic gesture meaningfulness $t(107) = 0.17$, $p = 0.86$, Cohen’s $d = 0.01$. Note that the group difference for nonsense gesture meaningfulness should be interpreted with caution due to the negligible effect size.

Figure 3a depicts the across-gesture correlations between the average meaningfulness rating assigned to each gesture by the ASD group and the rating assigned by the NA group. These ratings were extremely highly correlated for both iconic gestures $r(106) = 0.94$, $p < 0.001$ and nonsense gestures $r(52) = 0.91$, $p < 0.001$, indicating strong agreement between groups about the relative meaningfulness of certain gestures over others.

Response diversity

Main effects of both gesture category $F(1, 160) = 55.24$, $p < 0.001$, $\eta_p^2 = 0.26$ and group $F(1, 160) = 30.90$, $p < 0.001$, $\eta_p^2 = 0.17$ were present for response diversity (Fig. 2b).

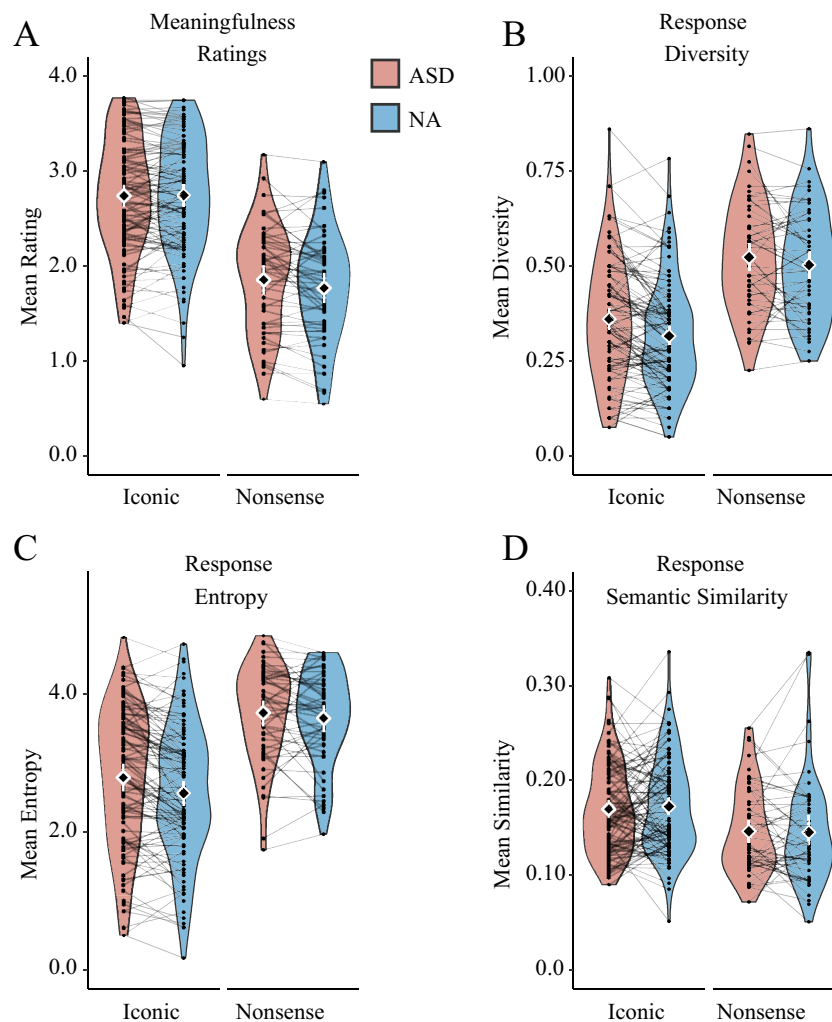


Fig. 2 Means and distributions of gesture norming metrics. *Note.* Violin plots depicting the means and distributions of **A** meaningfulness ratings, **B** response diversity, **C** response entropy, and **D** response semantic similarity, subdivided by gesture type (Iconic, Nonsense) and rater group (ASD, NA). *Small black dots* represent individual gestures and *y-axis* placement denotes the mean value

assigned to that gesture by ASD and NA raters. *Thin grey lines* connect each gesture's value from the ASD group with its corresponding value from the NA group. The *black diamonds outlined in white* represent condition means, and the *white lines* represent 95% confidence intervals

Specifically, nonsense gestures elicited a greater proportion of unique responses than did iconic gestures for both groups, and the response sets obtained from participants with ASD contained more unique labels on average relative to those obtained from participants without ASD. The interaction was non-significant $F(1, 160) = 2.79, p = 0.10, \eta_p^2 = 0.02$.

Figure 3b depicts the correlations between the number of distinct responses produced by the ASD group relative to the NA group for each gesture. As with meaningfulness ratings, across-gesture differences in response diversity between groups were highly correlated for both gesture types, $r(106) = 0.86, p < 0.001$ for iconic gestures, $r(52) = 0.81, p < 0.001$ for meaningless gestures.

Response entropy

Analyses of response entropy revealed main effects of both gesture category $F(1, 160) = 48.36, p < 0.001, \eta_p^2 = 0.23$ and group $F(1, 160) = 27.77, p < 0.001, \eta_p^2 = 0.15$ (Fig. 2c). Specifically, nonsense gestures elicited higher levels of response entropy relative to iconic gestures, and entropy was higher for response sets obtained from participants with ASD relative to those without ASD. These effects were qualified by a significant interaction, $F(1, 160) = 4.42, p = 0.04, \eta_p^2 = 0.03$. Follow-up *t* tests revealed that difference in entropy between the ASD and NA response sets was significant for iconic gestures $t(107) = 5.51, p < 0.001$, Cohen's $d = 0.23$,

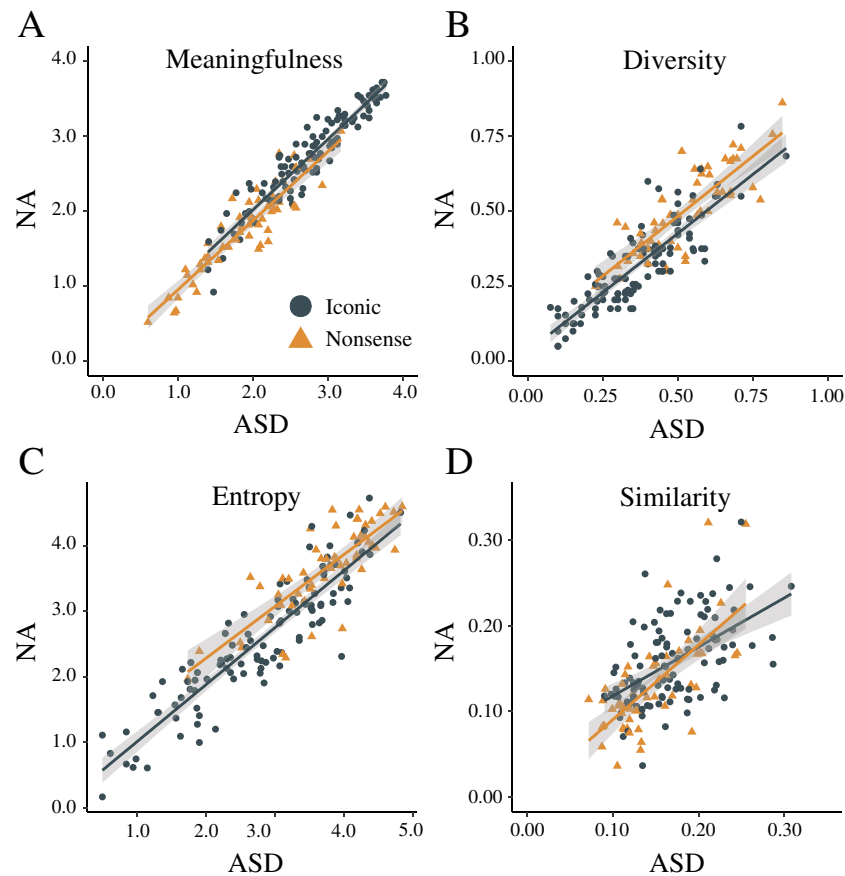


Fig. 3 Across-gesture correlations for gesture norming metrics between ASD and NA groups. *Note.* Scatterplots depicting across-gesture correlations between the ASD and NA groups for **A** meaningfulness ratings, **B** response diversity, **C** response entropy, and **D**

response semantic similarity, subdivided by gesture type (Iconic: *green circles*, Nonsense: *yellow triangles*). Each data point represents an individual gesture and denotes the mean value for each measure from the ASD group (x-axis) and NA group (y-axis)

but not for nonsense gestures $t(53) = 1.33, p = 0.19$, Cohen's $d = 0.11$.

As shown in Fig. 3c, strong and significant correlations across gestures were present between levels of response entropy produced by the ASD relative to the NA group for both iconic gestures $r(106) = 0.91, p < 0.001$ and nonsense gestures $r(52) = 0.81, p < 0.001$.

Response semantic similarity

A significant main effect of gesture category was present for response semantic similarity, $F(1, 160) = 12.09, p = 0.001, \eta_p^2 = 0.07$ (Fig. 2d). This effect reflected greater similarity among labels assigned to iconic relative to nonsense. The main effect of group was not significant, $F(1, 160) = 0.17, p = 0.68, \eta_p^2 = 0.00$, nor was the interaction, $F(1, 160) = 0.24, p = 0.63, \eta_p^2 = 0.00$.

Figure 3d depicts the correlation between the semantic similarity of the responses produced by the ASD group relative to the NA group for each gesture. The correlation was

significant for both iconic $r(106) = 0.54, p < 0.001$ and nonsense gestures $r(52) = 0.70, p < 0.001$.

Discussion

The goal of the present study was to provide a database of high-quality and well-characterized videos ranging in meaningfulness for use in gesture research. To our knowledge, only two sets of silent iconic gesture videos are currently available (Ortega and Özyürek, 2020; van Nispen et al., 2017), as well as one set that includes iconic gestures, emblems, and meaningless gestures (Agostini et al., 2019). Several features of the present database are unique from and complementary to these existing resources. First, each gesture video was carefully edited to be temporally uniform and precise, meaning that the videos are equal in length with the same amount of time before, during, and after each gesture. This uniformity is beneficial for methods such as event-related potentials (ERPs) that require precise attention to timing. Second, unlike other databases,

here actor visibility is limited to the neck down, which required us to avoid including gestures that incorporate the face or head (e.g., “applying lipstick”; Agostini et al., 2019). This lack of face visibility serves to eliminate influences from social cues such as facial expression or eye gaze and allows users of the videos to embed auditory speech without creating incongruity with respect to the actor’s mouth/facial movements.

Third, and perhaps most notably, each video was interpreted and rated for meaningfulness by groups of participants with and without a diagnosis of ASD. Research into the extent and nature of gesture processing difficulties in autistic children and adults has produced mixed results. Resolving these inconsistencies is of particular importance given that “deficits in understanding and use of gesture” is currently mentioned in the DSM-5 as one way to fulfill the diagnostic requirement of presenting difficulty with nonverbal communicative behavior (American Psychiatric Association, 2013). One potential source of variability in research outcomes is heterogeneity in the stimuli used across studies (see Agnostini et al., 2019 and Kandana Arachchige et al., 2021 for similar arguments). Well-characterized, openly available stimulus sets like the present database promise to accelerate research in this area by facilitating the ability for multiple research teams to use the same stimuli, and by providing stimuli for which people with and without ASD were equally represented in the norming process. We have also included ancillary measures, such as motion quantification and lemma age-of-acquisition, which may be particularly relevant to an autistic population.

The present study is the first to our knowledge to compare gesture naming and perceived gesture meaningfulness between people with and without ASD. The data contain several insights. First, use of the meaningfulness scale was highly similar between the groups. For gestures that were categorized *a priori* as iconic, no differences were found in perceived meaningfulness ratings across groups, and the elevated meaningfulness ratings in the ASD group for the nonsense gestures had a negligible effect size. Second, consensus meaningfulness ratings from each group were extremely highly correlated across gestures, as were measures of response diversity, response entropy, and response semantic similarity. This pattern of results suggests that the cues used to evaluate gestures for meaningfulness are comparable across groups, at least for stimuli such as these that contain only hand-based cues to meaning.

Group differences were present in the patterns of one-word names provided as stimulus labels. Relative to their non-autistic counterparts, autistic participants provided more unique gesture labels (i.e., greater response diversity) for gestures overall. Gesture labels provided by autistic

participants also had higher entropy scores, indicating a relative lack of dominance of certain responses over others. That said, this greater variability in word choice among autistic relative to non-autistic participants does not necessarily mean that the concepts evoked by the gestures were different or more variable among autistic participants. Indeed, although the overall number of responses was larger for the ASD group, analyses of the semantic similarity values among these responses revealed no differences between groups. As an illustrative example, for the iconic gesture *inserting* (Fig. 1c), the gesture labels *insert*, *enclose*, and *into* were given by participants in both groups. Autistic participants provided more than twice the number of unique responses for *inserting* as non-autistic participants, including *deposit*, *install*, and *sheath*. However, the meanings of these words are highly similar both to one another and to the gesture’s intended meaning.

It is helpful to consider the above results alongside research on forms of “unconventional language use” that have been documented in the autistic population, some of which are characterized by atypical word choices (for reviews, see Luyster et al., 2022 and Naigles & Tek, 2017). Dunn et al. (1996) found that children with ASD provide less prototypical examples of category members relative to both neurotypical and language-delayed children in a category fluency task in which they were asked to name animals and vehicles. This and similar findings have led to suggestions that ASD may be characterized by reduced lexical or semantic organization, sometimes combined with an advanced command of less frequent word forms (Naigles and Tek, 2017; Hilvert et al., 2019). Other forms of unconventional language use in ASD are social/pragmatic in nature, such as a tendency for autistic individuals to offer information that is more specific, technical, or detailed than is needed within a given discourse context (i.e., “pedantic speech”, De Villiers et al., 2007; Ghaziuddin & Gerstein, 1996). Accordingly, it is possible that the greater diversity of word choices provided by the ASD group during the naming task reflected different strategies evoked by our prompt (“If you had to choose one word to describe this gesture, what would it be?”). Non-autistic participants may have relied on lexical selection heuristics concerning communication efficiency and word accessibility (Koranda et al., 2022), whereas autistic raters may have been more likely to provide the most specific word that comes to mind. Future studies could test this possibility by varying the instructions given to the participant (e.g., “What is the very first word that this gesture brings to mind?”) and/or by using a debrief questionnaire to probe participants’ decision-making processes.

The online format of the current study presents a few limitations. No formal assessments of autism, language ability, or cognitive ability were conducted, and we relied on

self-report to categorize participants as either autistic or not autistic. Although we took several empirically supported precautions to minimize deception on the part of participants, some risk of intentional or unintentional misclassification remains. This caveat is particularly relevant to the null results presented here, including the null result regarding group differences in semantic similarity among gesture labels. This and other results implying a lack of group differences (e.g., the high across-gesture correlations among meaningfulness ratings) should be interpreted with caution until additional research is conducted in which diagnosis status and associated cognitive and communication differences are formally validated.

It is also important to note that the autistic individuals who participated in this study likely do not represent the full spectrum of ability and disability that can be associated with ASD. For example, all participants had the language and other cognitive skills necessary to follow instructions and complete the task, as well as to create a profile on Prolific. Accordingly, the generalizability of our results may be constrained to individuals on the autism spectrum who have lower support needs. We took steps to make our study accessible, such as using a self-paced and repetitive trial format and allowing the videos to be viewed multiple times. Our task also passed all Web Content Accessibility Guidelines (WCAG) checks with one exception: our optional demographics questionnaire included a matrix response table, which can be difficult for users with cognitive disability or low vision to complete. Nonetheless, future research that goes further in offering accommodations—for example, by including the option for a caregiver to assist with navigating the online platform, understanding instructions, and/or entering responses—would be beneficial to test the generalizability of these results and to determine the suitability of our stimulus database for research that involves autistic people with higher support needs.

The fact that we only recruited American-English speakers residing in the United States presents an additional constraint on generalizability. While our meaningful gestures were rated high in iconicity, the stimulus set contains a few gestures that could be appropriately described as *emblems* (McNeill, 1985). Emblematic gestures are those that are socially learned and often culture-specific, such as holding a thumbs-up to convey approval or satisfaction (Agostini et al, 2019). Differences among languages may also influence how a gesture is interpreted. For example, Ortega and Özyürek (2020) point out that in Dutch, an action and its accompanying tool are often incorporated into one word and provide the examples “*knippen*, ‘to cut with scissors’” and “*snijden*, ‘to cut with a knife’” (p. 56). Thus, the gestures included in this stimulus set may not be linguistically or culturally appropriate for populations outside of the U.S. or who primarily speak a language other than English.

Despite these limitations, our stimulus set is well positioned to accelerate research on a variety of research topics related to gesture processing, both with and without reference to ASD. Indeed, in a recent review paper, Kandana Arachchige et al. (2021) identified multiple areas of inconsistency across studies on how iconic gestures are integrated with speech that may stem from stimulus differences. For example, Habets et al. (2011) found evidence that incongruent gestures impair comprehension, but only when they are presented concurrently with speech or at stimulus onset asynchronies (SOAs) less than 360 ms. By contrast, Kelly et al. (2004) and Wu and Coulson (2007a) reported impairing effects of incongruent gestures at SOAs of 800 ms and >1000 ms, respectively. Kandana Arachchige et al. (2021) proposed that differences in gesture ambiguity across studies may account for discrepant findings, with asynchronous gestures primarily impacting language comprehension when they are relatively unambiguous. These authors also raised the possibility that speech-gesture integration impairment in ASD may differ in magnitude depending on whether the gestures provide information that is redundant versus complementary to the information provided by the speech (see also Dimitrova et al., 2017; Perrault et al, 2019). The detailed information provided about gesture meaningfulness and interpretation in the present stimulus database can aid in future research efforts to directly test these and other theoretically and clinically meaningful research questions related to gesture processing.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-023-02268-1>.

Acknowledgements We thank Cody Capps, Cristen Cooper, Dylan Howie, Brooke Montgomery, and Adrienne Wyble for assistance with video editing and data processing. A preliminary version of this research was presented at the 2021 Annual Meeting of the Cognitive Science Society.

Author’s Note Language—and even the diagnosis—of autism spectrum disorder has evolved over time and will likely continue to do so. Terminology once considered appropriate is now outdated (e.g., *low-functioning* and *high-functioning*). Once a separate diagnosis from Autism, “Asperger’s syndrome” was removed from the Diagnostic and Statistical Manual (DSM-V) in the American Psychiatric Association’s update in 2013. There is a growing movement to cease using the eponym altogether due to Hans Asperger’s involvement in the inhumane treatment and killing of sick and disabled children during World War II in Nazi-controlled Austria (Czech, 2018; Juntti, 2021; Reese, 2018). Recently, there has been a shift in the autism research field to move away from strictly using person-first language (i.e., *person with autism*) and toward identity-first language (i.e., *autistic person*; Botha et al., 2021; Vivanti, 2020). This is an ongoing conversation sparked by voices from the #ActuallyAutistic community and autistic self-advocates, whose members are proud of their neurodiversity and autistic identity (e.g., Leadbitter et al., 2021).

It is the authors’ sincere effort to use appropriate and inclusive language. We opted to use both identity-first and person-first language to respect the preference of autistic people and because part of our research interest concerns autism diagnosis status. Additionally, we recognize that some of the language from previous works cited here are no longer current (e.g., Asperger’s syndrome). Finally, in a good-faith

effort to understand autism and improve interventions and treatments, most autism research focuses on symptomology. However, many autistic people have voiced frustration over the tendency for researchers to focus on “a collection of deficits” (Bridge, 2018) while overlooking characteristics of worth and merit. We would like to emphasize that the idiosyncratic responses used to describe gestures by autistic participants in this study—and the development of a precocious vocabulary more broadly—are such features of autism that may be considered strengths rather than symptoms.

Funding This work was funded by a Louisiana Board of Regents Grant to HDL.

Data availability The data and materials pertaining to this study are available in the Open Science Framework repository at: <https://osf.io/a3pg7/>

Code availability The codes used to process and analyze these data are available at the link above.

Declarations

Conflicts of interest/Competing interests The authors have no conflicts of interest or competing interests to declare.

Ethics approval This study was approved by Louisiana State University’s Institutional Review Board.

Consent to participate All participants gave informed consent to participate. Participants’ identities were not disclosed to the researchers and the dataset contains no identifying information.

References

- Adornetti, I., Ferretti, F., Chiera, A., Waciewicz, S., Żywicznyński, P., Deriu, V., Marini, A., Magni, R., Casula, L., Vicari, S., & Valeri, G. (2019). Do children with autism spectrum disorders understand pantomimic events? *Frontiers in Psychology, 10*, 1382. <https://doi.org/10.3389/fpsyg.2019.01382>
- Agostini, B., Papeo, L., Galusca, C.-I., & Lingnau, A. (2019). A norming study of high-quality video clips of pantomimes, emblems, and meaningless gestures. *Behavior Research Methods, 51*(6), 2817–2826. <https://doi.org/10.3758/s13428-018-1159-8>
- Aldaqr, I., Schuwerk, T., Daum, M. M., Sodian, B., & Paulus, M. (2016). Sensitivity to communicative and non-communicative gestures in adolescents and adults with autism spectrum disorder: Saccadic and pupillary responses. *Experimental Brain Research, 234*(9), 2515–2527. <https://doi.org/10.1007/s00221-016-4656-y>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Association.
- Bartolo, A., Cubelli, R., Sala, S. D., Drei, S., & Marchetti, C. (2001). Double dissociation between meaningful and meaningless gesture reproduction in apraxia. *Cortex, 37*(5), 696–699. [https://doi.org/10.1016/S0010-9452\(08\)70617-8](https://doi.org/10.1016/S0010-9452(08)70617-8)
- Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy*. Academic Press.
- Beattie, G., & Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach. *Gesture, 1*(2), 129–149. <https://doi.org/10.1075/gest.1.2.03bea>
- Botha, M., Hanlon, J., Williams, G. L. (2021). Does language matter? Identity-first versus person-first language use in autism research: A response to Vivanti. *Journal of Autism and Developmental Disorders, 1–9*. <https://doi.org/10.1007/s10803-020-04858-w>
- Breckinridge Church, R., Garber, P., & Rogalski, K. (2007). The role of gesture in memory and social communication. *Gesture, 7*(2), 137–158. <https://doi.org/10.1075/gest.7.2.02bre>
- Bridge, C.L. (2018, August 21). “Actually, It’s Crooked”: Autism and Precocious Speech. Autistic Women & Nonbinary Network (AWN). Retrieved March of 2022 from <https://awnnetwork.org/actually-its-crooked-autism-and-precocious-speech/>
- Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2018). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. arXiv preprint arXiv:1812.08008. <https://doi.org/10.48550/arXiv.1812.08008>
- Chandler, J. J., & Paolacci, G. (2017). Lie for a dime: When most prescreening responses are honest but most study participants are impostors. *Social Psychological and Personality Science, 8*(5), 500–50. <https://doi.org/10.1177/1948550617698203>
- Clements, C., & Chawarska, K. (2010). Beyond pointing: Development of the “showing” gesture in children with autism spectrum disorder. *Yale Review of Undergraduate Research in Psychology, 2*, 1–11.
- Close, B., Zurbenko, I., Sun, M. (2020). *kza: Kolmogorov-Zurbenko-AdaptiveFilters*. R package version 4.1.0.1. Retrieved on July 21, 2023 from <https://CRAN.R-project.org/package=kza>
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language, 63*(4), 465–475. <https://doi.org/10.1016/j.jml.2010.07.002>
- Cook, J., Crane, L., Hull, L., Bourne, L., & Mandy, W. (2022). Self-reported camouflaging behaviours used by autistic adults during everyday social interactions. *Autism, 26*(2), 406–421. <https://doi.org/10.1177/13623613211026754>
- Cossu, G., Boria, S., Copioli, C., Bracceschi, R., Giuberti, V., Santelli, E., & Gallese, V. (2012). Motor representation of actions in children with autism. *PLoS ONE, 7*(9), e44779. <https://doi.org/10.1371/journal.pone.0044779>
- Czech, H. (2018). Hans Asperger, national socialism, and “race hygiene” in Nazi-era Vienna. *Molecular Autism, 9*(1), 1–43. <https://doi.org/10.1186/s13229-021-00433-x>
- Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin, 145*(8), 765–784. <https://doi.org/10.1037/bul0000202>
- Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology, 27*(3), 403–424. https://doi.org/10.1207/s15326942dn2703_6
- De Marchena, A., Kim, E. S., Bagdasarov, A., Parish-Morris, J., Maddox, B. B., Brodtkin, E. S., & Schultz, R. T. (2019). Atypicalities of gesture form and function in autistic adults. *Journal of Autism and Developmental Disorders, 49*, 1438–1454. <https://doi.org/10.1007/s10803-018-3829-x>
- de Marchena, A., & Eigsti, I.-M. (2010). Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism Research, 3*(6), 311–322. <https://doi.org/10.1002/aur.159>
- de Villiers, J., Fine, J., Ginsberg, G., Vaccarella, L., & Szatmari, P. (2007). Brief report: A scale for rating conversational impairment in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37*(7), 1375–1380. <https://doi.org/10.1007/s10803-006-0264-1>
- Dimitrova, N., Özçalışkan, Ş, & Adamson, L. B. (2017). Do verbal children with autism comprehend gesture as readily as typically developing children? *Journal of Autism and Developmental Disorders, 47*(10), 3267–3280. <https://doi.org/10.1007/s10803-017-3243-9>

- Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3), 445–454. <https://doi.org/10.1518/hfes.45.3.445.27258>
- Dunn, M., Gomes, H., & Sebastian, M. J. (1996). Prototypicality of responses of autistic, language disordered, and normal children in a word fluency task. *Child Neuropsychology*, 2(2), 99–108. <https://doi.org/10.1080/09297049608401355>
- Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, 18(2), 185–205. <https://doi.org/10.1080/09541440540000158>
- Fourie, E., Palser, E. R., Pokorny, J. J., Neff, M., & Rivera, S. M. (2020). Neural processing and production of gesture in children and adolescents with autism spectrum disorder. *Frontiers in Psychology*, 10, 3045. <https://doi.org/10.3389/fpsyg.2019.03045>
- Ghaziuddin, M., & Gerstein, L. (1996). Pedantic speaking style differentiates Asperger syndrome from high-functioning autism. *Journal of Autism and Developmental Disorders*, 26(6), 585–595. <https://doi.org/10.1007/BF02172348>
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3(11), 419–429. [https://doi.org/10.1016/S1364-6613\(99\)01397-2](https://doi.org/10.1016/S1364-6613(99)01397-2)
- Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience*, 23(8), 1845–1854. <https://doi.org/10.1162/jocn.2010.21462>
- Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137(2), 297–315. <https://doi.org/10.1037/a0022128>
- Hubbard, A. L., McNealy, K., Scott-Van Zeeland, A. A., Callan, D. E., Bookheimer, S. Y., & Dapretto, M. (2012). Altered integration of speech and gesture in children with autism spectrum disorders. *Brain and Behavior*, 2(5), 606–619. <https://doi.org/10.1002/brb3.81>
- Hurley, R. S. E., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. *Journal of Autism and Developmental Disorders*, 37(9), 1679–1690. <https://doi.org/10.1007/s10803-006-0299-3>
- Jongejan, B., Dalianis, H. (2009). Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - ACL-IJCNLP '09*, 1, 145. <https://doi.org/10.3115/1687878.1687900>
- Juntti, M. (July 21, 2021). It's time to stop calling Autism "Asperger's." *Fatherly*. Retrieved March of 2022 from <https://www.fatherly.com/health-science/aspergers-vs-autism-and-hans-asperger/>
- KandanaArachchige, K. G., SimoesLoureiro, I., Blekic, W., Rossignol, M., & Lefebvre, L. (2021). The role of iconic gestures in speech comprehension: An overview of various methodologies. *Frontiers in Psychology*, 12, 634074. <https://doi.org/10.3389/fpsyg.2021.634074>
- Kartalkanat, H., & Göksun, T. (2020). The effects of observing different gestures during storytelling on the recall of path and event information in 5-year-olds and adults. *Journal of Experimental Child Psychology*, 189, 104725. <https://doi.org/10.1016/j.jecp.2019.104725>
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, 89(1), 253–260. [https://doi.org/10.1016/S0093-934X\(03\)00335-3](https://doi.org/10.1016/S0093-934X(03)00335-3)
- Koranda, M. J., Zettersten, M., & MacDonald, M. C. (2022). Good-enough production: Selecting easier words instead of more accurate ones. *Psychological Science*, 33(9), 1440–1451. <https://doi.org/10.1177/09567976221089603>
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior research methods*, 44, 978–990. <https://doi.org/10.3758/s13428-012-0210-4>
- Lambrechts, A., Gaigg, S., Yarrow, K., Maras, K., & Fusaroli, R. (2014). Temporal dynamics of speech and gesture in Autism Spectrum Disorder. *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 349–353). IEEE. <https://doi.org/10.1109/CogInfoCom.2014.7020477>
- Lambrechts, A., Yarrow, K., Maras, K., & Gaigg, S. (2014). Impact of the temporal dynamics of speech and gesture on communication in autism spectrum disorder. *Procedia - Social and Behavioral Sciences*, 126, 214–215. <https://doi.org/10.1016/j.sbspro.2014.02.380>
- Leadbitter, K., Buckle, K. L., Ellis, C., Dekker, M. (2021). Autistic self-advocacy and the neurodiversity movement: Implications for autism early intervention research and practice. *Frontiers in Psychology*, 782. <https://doi.org/10.3389/fpsyg.2021.635690>
- Levantinou, E. I., Navarretta, C. (2016). An investigation of the effect of beat and iconic gestures on memory recall in L2 speakers. In: *Proceedings from the 3rd European symposium on multimodal communication* (Vol. 105, pp. 32–37).
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). The Autism Diagnostic Observation Schedule, Second Edition (ADOS-2). Western Psychological Services.
- Luyster, R. J., Zane, E., & Wisman Weil, L. (2022). Conventions for unconventional language: Revisiting a framework for spoken language features in autism. *Autism & Developmental Language Impairments*, 7, 239694152211054. <https://doi.org/10.1177/23969415221105472>
- Manwaring, S. S., Stevens, A. L., Mowdood, A., & Lackey, M. (2018). A scoping review of deictic gesture use in toddlers with or at-risk for autism spectrum disorder. *Autism & Developmental Language Impairments*, 3, 1–27. <https://doi.org/10.1177/2396941517751891>
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92(3), 350–371. <https://doi.org/10.1037/0033-295X.92.3.350>
- Mishra, A., Ceballos, V., Himmelwright, K., McCabe, S., & Scott, L. (2021). Gesture production in toddlers with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 51(5), 1658–1667. <https://doi.org/10.1007/s10803-020-04647-5>
- Naiges, L. R., & Tek, S. (2017). ‘Form is easy, meaning is hard’ revisited: (Re) characterizing the strengths and weaknesses of language in children with autism spectrum disorder. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(4), e1438. <https://doi.org/10.1002/wcs.1438>
- Ortega, G., & Özyürek, A. (2020). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods*, 52(1), 51–67. <https://doi.org/10.3758/s13428-019-01204-6>
- Özçalışkan, Ş., Adamson, L. B., & Dimitrova, N. (2016). Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism* 20(6), 754–763. <https://doi.org/10.1177/1362361315605921>
- Özyürek, A., & Woll, B. (2019). Language in the visual modality: Cospeech gesture and sign language. In *Human language: From genes and brain to behavior* (pp. 67–83). MIT Press.
- Pennington, J., Socher, R., Manning, C. (2014). Glove: Global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- Perrault, A., Chaby, L., Bigouret, F., Oppetit, A., Cohen, D., Plaza, M., & Xavier, J. (2019). Comprehension of conventional gestures in typical children, children with autism spectrum disorders and children with language disorders. *Neuropsychiatrie de l'Enfance et de l'Adolescence*, 67(1), 1–9. <https://doi.org/10.1016/j.neuinf.2018.03.002>
- Reese, H. (2018, May 22). The disturbing history of Dr. Asperger and the movement to reframe the syndrome. *Vox*. Retrieved March

- 2022 from <https://www.vox.com/conversations/2018/5/22/17377766/asperger-nazi-rename-syndrome>
- Rinker, T. W. (2018a). *lexicon: Lexicon Data*. R package version 1.2.1, retrieved on Apr. 13, 2022 from <http://github.com/trinker/lexicon>.
- Rinker, T. W. (2018b). *textstem: Tools for stemming and lemmatizing text*. R package version 0.1.4, Buffalo, New York. Retrieved on Apr. 13, 2022 from <http://github.com/trinker/textstem>.
- Selivanov, D., Bickel, M., Wang, Q. (2022) *text2vec: Modern text mining framework for R*. R package version 0.6.3, retrieved on Dec. 1, 2022 from <https://CRAN.R-project.org/package=text2vec>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Silverman, L. B., Eigsti, I.-M., & Bennetto, L. (2017). I taw i taw a puddy tat: Gestures in canary row narrations by high-functioning youth with autism spectrum disorder: Gesture production in ASD. *Autism Research*, 10(8), 1353–1363. <https://doi.org/10.1002/aur.1785>
- So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, 27(5), 665–681. <https://doi.org/10.1080/01690965.2011.573220>
- Stieglitz Ham, H., Bartolo, A., Corley, M., Rajendran, G., Szabo, A., & Swanson, S. (2011). Exploring the relationship between gestural recognition and imitation: Evidence of dyspraxia in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 41(1), 1–12. <https://doi.org/10.1007/s10803-010-1011-1>
- Straube, B., Meyer, L., Green, A., & Kircher, T. (2014). Semantic relation vs. surprise: The differential effects of related and unrelated coverbal gestures on neural encoding and subsequent recognition. *Brain Research*, 1567, 42–56. <https://doi.org/10.1016/j.brainres.2014.04.012>
- Thompson, L. A., Driscoll, D., & Markson, L. (1998). Memory for visual-spoken language in children and adults. *Journal of Nonverbal Behavior*, 22(3), 167–187. <https://doi.org/10.1023/A:1022914521401>
- Trettenbrein, P. C., Zaccarella, E. (2021). Controlling video stimuli in sign language and gesture research: The OpenPoseR package for analyzing OpenPose motion-tracking data in R. *Frontiers in Psychology*, 12, <https://doi.org/10.3389/fpsyg.2021.628728>
- Trujillo, J. P., Özyürek, A., Kan, C. C., Sheftel-Simanova, I., & Bekkering, H. (2021). Differences in the production and perception of communicative kinematics in autism. *Autism Research*, 14(12), 2640–2653. <https://doi.org/10.1002/aur.2611>
- van Nispen, K., van de Sandt-Koenderman, W., Mieke, E., & Kraemer, E. (2017). Production and comprehension of pantomimes used to depict objects. *Frontiers in Psychology*, 8, 1095. <https://doi.org/10.3389/fpsyg.2017.01095>
- Vivanti, G. (2020). Ask the editor: What is the most appropriate way to talk about individuals with a diagnosis of autism? *Journal of Autism and Developmental Disorders*, 50(2), 691–693. <https://doi.org/10.1007/s10803-019-04280-x>
- Vivanti, G., Nadig, A., Ozonoff, S., & Rogers, S. J. (2008). What do children with autism attend to during imitation tasks? *Journal of Experimental Child Psychology*, 101(3), 186–205. <https://doi.org/10.1016/j.jecp.2008.04.008>
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1109/TIP.2003.819861>
- Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., & Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: A retrospective home video study. *American Journal of Speech-Language Pathology*, 22(1), 25–39. [https://doi.org/10.1044/1058-0360\(2012/11-0145\)](https://doi.org/10.1044/1058-0360(2012/11-0145))
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6), 654–667. <https://doi.org/10.1111/j.1469-8986.2005.00356.x>
- Wu, Y. C., & Coulson, S. (2007). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*, 14(1), 57–63. <https://doi.org/10.3758/BF03194028>
- Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, 101(3), 234–245. <https://doi.org/10.1016/j.bandl.2006.12.003>

Open Practices Statement The gesture videos and corresponding norming data, along with data processing and analysis scripts, are available at <https://osf.io/a3pg7/>. For questions concerning the database or access to the materials, contact Brianna E. Cairney at bcainr2@lsu.edu. This study was not pre-registered.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.