



OPEN

Identifying the signature of prospective motor control in children with autism

Andrea Cavallo^{1,2}, Luca Romeo^{3,4}, Caterina Ansuini¹, Francesca Battaglia^{1,5}, Lino Nobili^{5,6}, Massimiliano Pontil⁴, Stefano Panzeri⁷ & Cristina Becchio^{1✉}

Failure to develop prospective motor control has been proposed to be a core phenotypic marker of autism spectrum disorders (ASD). However, whether genuine differences in prospective motor control permit discriminating between ASD and non-ASD profiles over and above individual differences in motor output remains unclear. Here, we combined high precision measures of hand movement kinematics and rigorous machine learning analyses to determine the true power of prospective movement data to differentiate children with autism and typically developing children. Our results show that while movement is unique to each individual, variations in the kinematic patterning of sequential grasping movements genuinely differentiate children with autism from typically developing children. These findings provide quantitative evidence for a prospective motor control impairment in autism and indicate the potential to draw inferences about autism on the basis of movement kinematics.

Much autism research has focused on the social, communication and cognitive difficulties associated with the condition. However, recent years have seen an increasing interest in the motor side of autism spectrum disorders (ASD)¹, with several researchers going as far as proposing that impairments in the prospective control of movements^{2,3} may be predictive of ASD and may even underlie some of the core features of ASD⁴. Despite this enthusiasm, relatively little quantitative information is available about prospective movement alterations in ASD and questions remain regarding whether (and to what extent) kinematic patterning permit differentiating ASD and non-ASD profiles over and above individual differences in motor output⁵.

Recently, researchers have started to address these questions by applying machine learning as a tool for multivariate analysis with the goal of finding inherent patterns in kinematic data. In a typical application, a machine learning classifier is trained to distinguish ASD and non-ASD movement patterns on one part of a data set. Then, the classifier is tested on the remaining data. This results in a certain fraction of accurate classifications. If the classification accuracy lies significantly above the level expected by chance (e.g., 0.50), it can be concluded that a difference between classes exists. Studies applying this approach to prospective movement data suggest that patterns related to ASD can be classified with near-perfect accuracy^{6,7}. The caveats inherent to applications of machine learning methods for identifying group differences in movement data, however, are rarely considered⁸.

A first caveat relates to confound variables, that is, variables that in the selected sample have an association with diagnosis group, but are uninteresting from a clinical perspective⁹. Consider, for example, movement variations related to age. In a sample in which ASD children, say, are younger than typically developing (TD) children⁷, the classifier is not only learning movement features discriminating ASD and TD movement, but also age-related features. This can lead to overestimate the information that is computationally available to differentiate ASD and TD movements.

A second caveat relates to individual variations in motor outputs. Individuals show variations in motor outputs that are both consistent within a given individual and differ from one individual to another^{10–13}. When repeated measures from the same individual are randomly assigned to training and testing as in common machine learning

¹Cognition, Motion and Neuroscience Laboratory, Center for Human Technologies, Istituto Italiano di Tecnologia, Genoa, Italy. ²Department of Psychology, Università degli Studi di Torino, Turin, Italy. ³Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy. ⁴Computational Statistics and Machine Learning Laboratory, Center for Human Technologies, Istituto Italiano di Tecnologia, Genoa, Italy. ⁵Child Neuropsychiatric Unit, IRCCS Istituto G. Gaslini, Genoa, Italy. ⁶DINOGLMI Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics and Maternal and Children's Sciences, Università degli Studi di Genova, Genoa, Italy. ⁷Neural Computational Laboratory, Center for Human Technologies, Istituto Italiano di Tecnologia, Genoa, Italy. ✉email: cristina.becchio@iit.it

approaches using record-wise cross-validation⁶, algorithms can learn the association between individuals' movement signature and their diagnostic label. This can lead to identity-confounded predictions, where the easier task of identity classification de facto circumvents and replaces the harder task of diagnostic classification¹⁴.

Both these caveats are addressed in the present study. In a sample of cases accurately matched on age, gender, full-scale IQ, and stature, we use machine learning methods as a tool to determine the true power of prospective movement data to differentiate ASD and TD group profiles and predict the group of a previously unseen individual. We do so by contrasting record-wise data splits, where repeated measurements of each individual are assigned to both the training set and the test set, with subject-wise data splits, where any source of identity confound is neutralized by the assignment of all measurements of each subject to either the training set or the test set¹⁴ (see “Methods” section for details). Our results show that although movement is unique to each individual, differences in prospective motor control genuinely differentiate ASD from TD movement data.

Results

Data for this study consisted of 1600 movements recorded from 20 ASD and 20 TD children reaching towards and grasping a bottle with one of three prospective intentions: place the bottle into a box (grasp-to-place), pour some water (grasp-to-pour) or pass the bottle to another person (grasp-to-pass; see “Methods” section). Visually, movement traces showed high variability across individuals (Fig. 1B–F) and trials (Supplementary Fig. S1), with variations across individuals greatly exceeding variations between groups (ASD, TD) in all extracted variables.

Intention classification. To quantify prospective modulation, we first attempted to classify intention, operationalized as onward action (place, pour, pass), from reach-to-grasp parameters, separately for TD and ASD data. Cross-validation using record-wise splits revealed that, in both groups, kinematics predicted intention above chance (TD mean \pm SEM = 0.64 ± 0.02 ; $p < 0.001$, permutation test; ASD mean \pm SEM = 0.57 ± 0.02 ; $p < 0.001$, permutation test). In line with¹⁵, prospective modulation was less pronounced in ASD movement profiles compared to TD movement profiles as indicated by a lower classification accuracy ($p = 0.039$, sign test).

Group classification. After establishing reduced prospective modulation in ASD compared to TD data, we next sought to determine the power of prospective movement data to differentiate ASD and TD movement profiles. Analysis of simulated¹⁴ and empirical data¹⁶ suggest that classifiers trained and evaluated using record-wise data splits can pick confounding relationship between identity and group and so produce inflated accuracies. To address this concern in our data, we contrasted record-wise data splits, where repeated measurements from the same individual are assigned to both the training set and the test set, with subject-wise data splits, where measures of each subject are assigned to either the training set or test set, neutralizing any potential identity confounding¹⁴. Classifiers trained and evaluated using record-wise splits achieved perfect accuracy in discriminating ASD versus TD (accuracy mean \pm SEM = 1.00 ± 0.00 ; sensitivity mean \pm SEM = 1.00 ± 0.00 ; specificity mean \pm SEM = 1.00 ± 0.00). As a control, we repeated record-wise splitting over 40 folds and found again that all cases were correctly classified (accuracy mean \pm SEM = 1.00 ± 0.00 ; sensitivity mean \pm SEM = 1.00 ± 0.00 ; specificity mean \pm SEM = 1.00 ± 0.00). Classification accuracies of classifiers trained with subject-wise data splits (accuracy mean \pm SEM = 0.75 ± 0.06 ; sensitivity mean \pm SEM = 0.75 ± 0.09 ; specificity mean \pm SEM = 0.75 ± 0.08) were significantly lower ($p = 0.002$, sign test), although still significantly above chance ($p = 0.010$, permutation test) (Fig. 2A). We obtained similar results using SVM-LASSO and RF (Supplementary Fig. S2).

To evaluate the influence of the training set size on classification accuracy, we considered four sets of classifiers in which the training set was composed of 4, 10, 20 or 39 subjects. The accuracy of classifiers built using record-wise splits was already perfect with 4 subjects and remained so through the four sets (Fig. 2D), raising the concern that record-wise classifiers fit to the idiosyncratic characteristics of the training sample^{17,18}. In contrast, the accuracy of classifiers built using subject-wise data splits was at chance with 4 and 10 subjects, increased to 0.61 with 20 subjects and continued to increase with 39 subjects, suggesting that subject-wise classifiers learned group information from new examples.

Quantification of identity confounding. The above results are compatible with the notion that identity confounding artificially inflates classification accuracies using record-wise data splits. To confirm the presence of identity information, as a first step, we tested whether group could be identified based on identity. In this analysis, we trained classifiers with record-wise splits to predict the identity of participants rather than the group. Consistent with the possibility that the power of record-wise classifiers is based on identity information, this analysis yielded perfect identity classification (accuracy mean \pm SEM = 1.00 ± 0.00 ; sensitivity mean \pm SEM = 1.00 ± 0.00 ; specificity mean \pm SEM = 1.00 ± 0.00).

To more directly test whether record-wise classifiers learn identity information rather than group information, we next randomly permuted the diagnostic labels of each subject as a block, so that all records of a given subject were assigned either ASD or TD during the permutation process^{16,19}. The motivation for this relabelling scheme is that it preserves the confounding association between group and identity, while breaking the relationship between movement data and group. Because group information is removed, permutation null distributions located on average above the 0.5 chance level indicate identity confounding¹⁹. Conforming an identity confounding, the null distribution of classification accuracies obtained from record-wise data splits had all data distributed above the 0.5 chance level (Fig. 2B). To assess whether classifiers were learning group information in addition to identity information, we compared the location of the record-wise classification accuracy computed using the original (unpermuted) data relative to the permuted data. Only if classification accuracy computed using the original data is higher than that of permuted data, it can be concluded that the classifier is learning group information. Record-wise accuracy computed using the original data did not differ from the permutation null

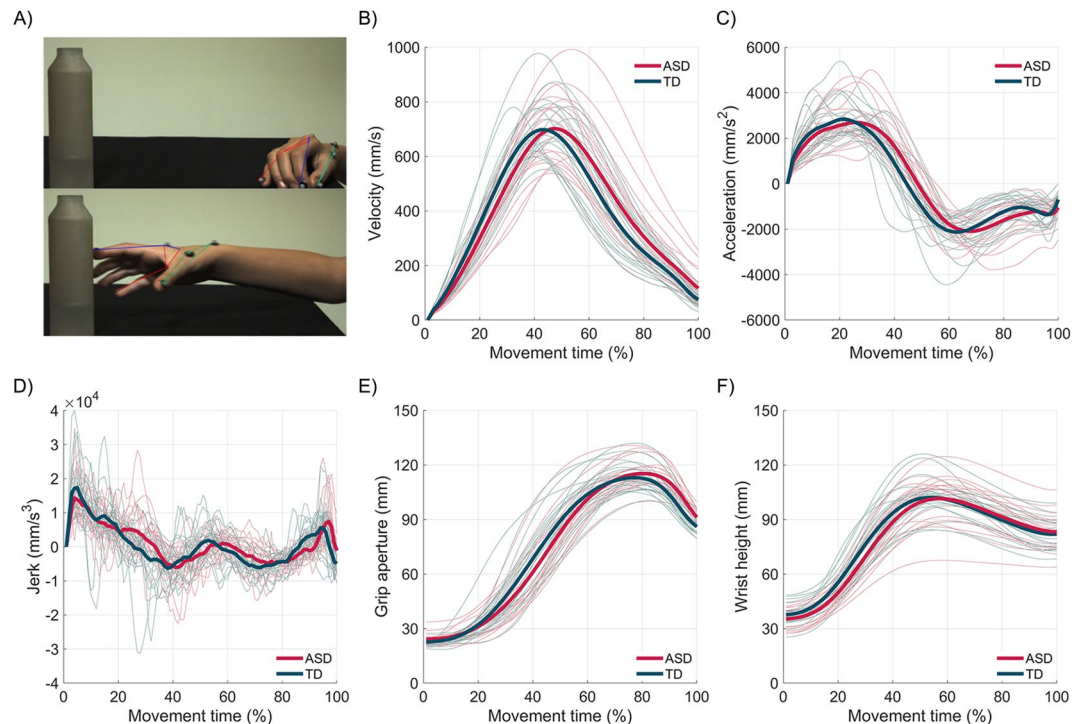


Figure 1. (A) Image sequence of a reach-to-grasp movement with hand model overlay. (B–F) Kinematic profiles of reach-to-grasp movements performed with three prospective intentions (to place, to pour, to pass). Wrist velocity (B), wrist acceleration (C), wrist jerk (D), grip aperture (E) and wrist height (F) were extracted. Thick lines represent group average (red = ASD group, green = TD group). Thin lines represent trial-average of individual participants.

distribution ($p=0.170$). In contrast to record-wise data splits, the null distribution generated by splitting data in a subject-wise fashion was located around 0.5 (Fig. 2C). Also, subject-wise accuracy computed using the original data was significantly higher than the adjusted null distribution ($p=0.01$, permutation test, Fig. 1D). Overall, these results support the conclusion that classifiers trained subject-wise, but not classifiers trained record-wise, learned differences between ASD and TD movements.

Generalizability to unseen cases. If a classifier is learning group information, its predictive power should generalize to unseen cases. In contrast, if a classifier is learning identity information, its accuracy should drop on unseen cases. Based on these predictions, we compared the ability of classifiers trained record-wise and subject-wise to predict group for unseen cases. To quantify the ability of classifiers trained record-wise to predict group for an unseen individual, we used a hybrid cross-validation procedure²⁰ in which hyper-parameters were tuned record-wise in the inner loop (model validation), whilst model performance was recursively tested subject-wise in the outer loop (model testing) on a new individual. Under this procedure, classifiers tuned record-wise (accuracy mean \pm SEM = 0.58 ± 0.08 ; sensitivity mean \pm SEM = 0.55 ± 0.12 ; specificity mean \pm SEM = 0.60 ± 0.10) performed at chance level ($p=0.110$, permutation test), significantly worse than classifiers tuned subject-wise at classifying unseen cases ($p=0.023$, sign test; see Fig. 2E). Figure 2F,G provide an example of classification of the movements of a new, unseen ASD participant using subject-wise (Fig. 2F) and record-wise (Fig. 2G) tuning. Data are visualized in a reduced kinematic space spanning only two kinematic features, velocity at 80% of movement duration and grip aperture at 90% of movement duration. Compared to the number of movements assigned to the correct group (ASD) using subject-wise data splits (30 out of 40; Fig. 2F), the number of correctly classified movements substantially decreases using record-wise data splits (17 out of 40; Fig. 2G). As a result, the participant's group (ASD) is correctly classified using subject-wise tuning, but not record-wise tuning.

Intention-specific contribution to group classification. Subject-wise data splits indicate that prospective movement data contain group discriminating information that generalizes to unseen cases. To examine how such information is encoded in specific intention profiles, we attempted to predict group separately from kinematic parameters of grasp-to-place, grasp-to-pour and grasp-to-pass movements. Supplementary Fig. S3 visualizes individual movement profiles graphed by intention. Separating intentions revealed that group differences were predominantly located in grasp-to-pass and grasp-to-pour movements. Characteristically, ASD grasp-to-pass showed later temporal setting of peak wrist velocity, peak wrist acceleration and peak wrist deceleration compared to TD grasp-to-pass. ASD grasp-to-pour also showed later peak wrist velocity and peak wrist deceleration compared to TD grasp-to-pour. Consistent with the visual impression gained from Supplementary Fig. S3, classification of group was most accurate from grasp-to-pass movements (accuracy

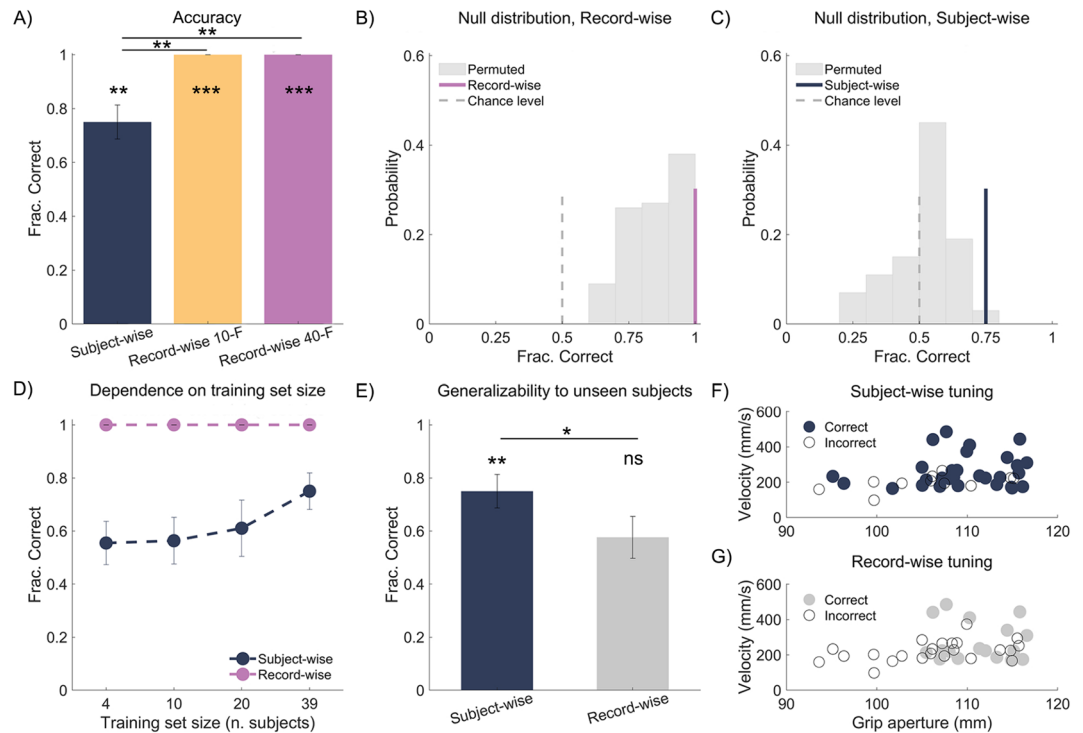


Figure 2. (A) Classification accuracy of group (ASD, TD) using subject-wise, record-wise 10 folds cross-validation and record-wise 40 folds cross-validation. Histograms represent mean \pm SEM. (B,C) Permutation null distribution generated by splitting data in a record-wise fashion (B) or subject-wise fashion (C). The permutation null distribution is represented by the grey histograms. The purple line in (B) represents the observed record-wise accuracy. The dark blue line in (C) represents the observed subject-wise accuracy. The dashed grey line indicates the 0.5 chance level. (D) Group classification accuracy as a function of sample size computed using record-wise and subject-wise cross-validation. (E) Group classification accuracy of unseen subjects obtained using subject-wise hyper-parameter tuning and record-wise hyper-parameter tuning. (F,G) Movement data colored according to group classification of one exemplar unseen subject with ASD using subject-wise hyper-parameter tuning (F) and record-wise hyper-parameter tuning (G). Data are visualized as a scatterplot of velocity at 80% of movement duration and grip aperture at 90% of movement durations.

mean \pm SEM = 0.83 ± 0.05 ; sensitivity mean \pm SEM = 0.80 ± 0.08 ; specificity mean \pm SEM = 0.85 ± 0.08), followed by grasp-to-pour movements (accuracy mean \pm SEM = 0.68 ± 0.07 ; sensitivity mean \pm SEM = 0.60 ± 0.11 ; specificity mean \pm SEM = 0.75 ± 0.09) and grasp-to-place movements (accuracy mean \pm SEM = 0.60 ± 0.07 ; sensitivity mean \pm SEM = 0.65 ± 0.12 ; specificity = 0.55 ± 0.12). Classification accuracy was above chance for grasp-to-pass ($p < 0.001$) and grasp-to-pour ($p = 0.02$) movements, but not for grasp-to-place movements ($p = 0.380$). This indicates that grasp-to-pass and grasp-to-pour movements, but not grasp-to-place movements, individually contributed group discriminating information.

Discussion

Developments in machine learning and the increasing availability of motion tracking devices have brought movement to the forefront of autism research, diagnosis and treatment^{4,21–23}. However, questions remain regarding whether and to what extent kinematic patterning permit differentiating ASD and non-ASD movement profiles⁵.

Here, we assessed the potential of multivariate prospective motor control data to discriminate movements performed by ASD children from TD children in a sample of cases matched on age, gender, full-scale IQ, and stature. By comparing record-wise and subject-wise cross-validation approaches in combination with different machine learning classifiers, we demonstrate that record-wise data splitting leads to an optimistic bias in accuracy. We show that such bias is inadvertently introduced in record-wise computations by individual movement characteristics. Individuals performing prospective grasps exhibit repeatable aspects of their movement signature across trials¹⁰. When data collected from the same individual are assigned to both the training and the test sets (as in record-wise cross-validation), models learn the idiosyncratic movement signature of individual participants, rather than the signature that separates ASD and TD movements. The upshot is that the apparent perfect classification accuracy achieved by classifiers trained record-wise drops to chance level on unseen cases.

Whit subject-wise cross-validation, the accuracy of prospective movement data in predicting group decreases—the best classifier achieving 0.75 classification accuracy—but remains statistically significant. Because data from each individual are assigned to either the training or the test set, subject-wise accuracies represent the true accuracy of prospective movement data in predicting group.

Separating data into intention-specific profiles revealed that group information is predominantly encoded in grasp-to-pass prospective movements. Unlike individual sequential actions such as grasp-to-place, passing an object requires the coordination—in space and time—between two partners^{24–26}. One interpretation of this result is that ASD atypical grasp-to-pass profiles relate to the social nature of the task. However, another possibility is that atypical grasp-to-pass profiles relate to the semi-predictable nature of the passing task²⁷. During passing, each partner may have a basic internal model regarding the evolution of the passing movement. However, given that control by one partner cannot be fully predicted by the other partner, there will necessarily be unpredictable aspects to the task²⁸. We propose that the later temporal setting of key landmarks in ASD grasp-to-pass movements may reflect reduced motor anticipation under conditions of uncertainty. A similar explanation may hold for grasp-to-pour movements, which have been shown to involve demanding probabilistic inferences about fluid dynamics over short time scales²⁹. This account makes the prediction, testable in future studies, that ASD and TD movement profiles become more distinct under conditions of increased volatility³⁰.

In our study, we chose to measure a homogeneous sample, carefully selected based on a specific set of inclusion and exclusion criteria to rule out a number of potential confounders. However, models based on homogeneous samples are less likely to generalize to real-life clinical settings, where patient groups are highly heterogeneous³¹. To dissect the heterogeneity of ASD, future studies should compare prospective motor control in ASD children with and without intellectual disability, and in ASD children classified as having comorbid attention and/or motor impairments. Such knowledge could help identifying ASD subtypes most likely to respond to specific interventions and developing personalized treatments.

Future research should also evaluate the specificity and sensitivity of ASD movement profiles relative to other neurodevelopmental disorders, such as attention deficit hyperactivity disorder and developmental coordination disorder. Testing specificity over multiple alternatives is important to establish whether prospective motor impairments in ASD could represent a putative endophenotype with potential to elucidate pathophysiology of ASD and facilitate early diagnosis⁵.

Finally, it will be important for future studies to assess the potential of prospective data to identify ASD in larger samples. In this regard, the steady increase in prediction accuracy observed using subject-wise data splits as a function of the size of training set—far from plateauing with 20 subjects per class—shows promise, suggesting that prediction of unseen cases could be improved when training models on larger samples.

Conclusions

The study reports on a promising advance in objectively characterizing movement kinematics in ASD. Our approach enabled us to isolate the information that genuinely differentiates ASD and TD movement profiles and to quantify the identity confounding inherent to record-wise cross-validation. Combined, our findings invite a reassessment of recent conclusions about the diagnostic power of prospective movement data. The finding that prospective movement data contain identity information, but also genuine group information highlights the utility of machine learning methods as a tool for developing robust single-subject predictor to accurately differentiate autism and/or subtypes of autism.

Methods

Participants. We report results from 20 ASD children (18 males) without accompanying intellectual impairment and 20 typically developing children (16 males). Groups were matched for age (TD mean \pm SD = 9.5 \pm 1.5 years, months; ASD mean \pm SD = 9.8 \pm 1.5 years, months; $t_{38} = -0.665$, $p = 0.510$), stature (TD mean \pm SD = 140.4 \pm 8.1 cm; ASD mean \pm SD = 137.7 \pm 9.1 cm; $t_{38} = 0.988$, $p = 0.329$) gender, and Full Scale IQ as measured by the Wechsler Scale of Intelligence (WISC IV)³² (TD mean \pm SD = 102.8 \pm 9.4; ASD mean \pm SD = 98.5 \pm 11.1; $t_{38} = 1.325$, $p = 0.193$). Children with ASD were diagnosed according to DSM-5 criteria³³. The Autism Diagnostic Observation Scale (ADOS-2)³⁴ and Autism Diagnostic Interview-Revised (ADI-R)³⁵ were administered by two experienced professionals. All children had normal or corrected-to-normal vision and were screened for exclusion criteria (pharmacological treatment, epilepsy, and any other neurological and psychiatric conditions). Both ASD and TD groups were assessed for executive functions abilities by means of the Tower of London (TOL) test³⁶. This test revealed no significant differences between TD and ASD children (TD mean \pm SD = 29.35 \pm 3.54; ASD mean \pm SD = 29.35 \pm 2.80; $t_{38} = 0$, $p = 0.999$). All but two of the children (one in the ASD group and one in the TD group) were right-handed according to the Edinburgh Handedness Inventory³⁷. Written informed consent was obtained from the parents of the children prior to participation in the experiment. The research protocol was approved by the local ethics committee (Comitato Etico Regionale Liguria) and was in accordance with the principles of the revised Helsinki Declaration (2013)³⁸.

Experimental design. Children were seated on a height-adjustable chair with their right elbow and wrist resting on a table (height = 64 cm; length = 100 cm; width = 60 cm). A plastic bottle filled with water (base diameter = 5 cm; height = 18 cm; weight = 225 g) was positioned on the table at a distance of 44 cm from children's midline. Children were asked to reach for and grasp the bottle to complete three sequential manipulation tasks: place the bottle into a box (grasp-to-place), pour some water into a glass (grasp-to-pour), or pass the bottle to a co-actor (grasp-to-pass), who would then either place the bottle into the box or pour some water. In each trial, children were asked to perform at a natural speed after an auditory tone. Each child completed 4 blocks of 12 trials (1 grasp-to-place block, 1 grasp-to-pour block, and 2 grasp-to-pass blocks). The order of blocks was pseudo-randomized across participants. The experiment lasted about 30 min, with a 2-min pause at the end of each block. Throughout the entire experimental session, the same female experimenter (co-actor) sat at the opposite side of the table and interacted with the child.

Kinematics recording and data processing. A near-infrared camera motion capture system equipped with six cameras (frame rate = 100 Hz; Bonita Vicon Motion Systems Ltd, Oxford, UK) was used to track and record the reach-to-grasp kinematics. The child's right hand was outfitted with 8 retro-reflective hemispheric markers (6.5 mm in diameter) placed on the metacarpal joint and the tip of the index and the little finger, the trapezium bone of the thumb, the radial aspect of the wrist and the centre of the hand dorsum (Fig. 1A). After data collection, each trial was individually inspected for correct marker identification and then run through a 6 Hz low-pass Butterworth filter¹⁵. Trials in which the quality of markers reconstruction was poor were discarded from the dataset and not considered for further analyses. To equate the total number of trials between blocks and participants, the final dataset consisted of 10 trials per block per participant, for a total of 1600 movements. A custom software written in Matlab (MathWorks, Natick, MA) was used to compute wrist velocity, wrist acceleration, wrist jerk, wrist height, and grip aperture. Each variable was computed at intervals of 10% of movement duration, from reach onset to reach offset (see Supplementary Table S1).

Classification analyses. *Intention classification.* We computed classification of onward action (place, pour, pass) from reach-to-grasp parameters based on a Support Vector Machine (SVM) model, separately for TD and ASD movement data. The SVM used Gaussian kernel (SVM-G) to compute the hyperplane that best separated grasp-to-place, grasp-to-pour and grasp-to-pass trials in each diagnosis group. To equate the number of trials across action classes, for grasp-to-pass we randomly selected 5 trials per block. For each diagnosis group, the dataset thus comprised 600 reach-to-grasp movements (20 participants \times 10 trials \times 3 actions). Models were trained, validated and tested using a record-wise cross-validation data split (with data split into 10 folds, each fold containing 1 movement for each intention for each participant). Classification accuracy was computed as the fraction of correctly classified trials for each cross-validation iteration.

Group classification. For this analysis, we used all 1600 reach-to-grasp movements (40 participants \times 10 trials \times 4 blocks). We computed group (ASD, TD) from reach-to-grasp parameters based on a SVM-G model. For comparison, we also computed group based on SVM regularized with least absolute shrinkage and selection operator (SVM-LASSO) and Random Forest (RF). Models were evaluated using record-wise cross-validation and subject-wise cross validation. For record-wise cross-validation, we split data into 10 (40) folds so that each fold contained 4 (1) movements from each subject. For subject-wise cross-validation, we split data by subjects such that training and test folds contained records from different subjects. In both methods, hyper-parameters were tuned recursively on all but one fold of the training set and tested on the remaining fold. For each subject, group was determined based on the averaged posterior scores of the movements of that subject. Classification accuracy was computed as the fraction of correctly classified subjects. Sensitivity—the fraction of ASD cases correctly classified as ASD in reference to all ASD cases—and specificity—the fraction of TD cases correctly classified as TD in reference to all TD cases—were also computed as measures of classifier performance.

Intention-specific contribution to group classification. To evaluate the diagnostic power of specific action sequences, we also computed group separately from grasp-to-place, grasp-to-pour and grasp-to-pass movements based on three separate SVM-G models. Models were validated and tested using subject-wise cross validation.

Statistics of model performance. In all analyses, we assessed model performance using classification accuracy. We evaluated the significance of classification accuracy against chance with permutation statistics. The chance-level null-hypothesis distribution of these statistics was created by computing classification accuracy after randomly permuting the trial label of each individual trial for the record-wise calculations and the trial label of all trials of a given subject as block for the subject-wise calculations (100 random permutations). We evaluated the significance of differences in how accurately different models classified intentions (intention classification) or group (group classification) using two-tailed paired-sample sign tests. We used non-parametric tests because classification accuracies do not follow a Gaussian distribution. Standard error of the mean (SEM) were computed by bootstrapping for reporting, but their value was not used in the computation of the non-parametric statistics of model performance.

Notes In all figures, * indicates $p < 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$. Following standard notations, asterisks above bars indicate above chance significance, asterisks above connecting lines indicate comparison significance.

Community involvement. We consulted clinical practitioners from the local hospital on how to create an enabling environment and plan experimental procedures to maximise the engagement and enjoyment of the children. Their feedback was incorporated in the study design. We shared the aggregate study results with children and their families in a 1-day workshop organized by the research team in partnership with clinical practitioners.

Data availability

The data supporting the main findings of the current study are available from the corresponding author on reasonable request. The code supporting the main findings is based on public available tools as detailed in “Methods” section. Custom functions inputting data to toolboxes will be made available by the corresponding author upon reasonable request.

Received: 8 December 2020; Accepted: 19 January 2021

Published online: 04 February 2021

References

1. Sacrey, L.-A.R., Germani, T., Bryson, S. E. & Zwaigenbaum, L. Reaching and grasping in autism spectrum disorder: a review of recent literature. *Front. Neurol.* **5**, 6. <https://doi.org/10.3389/fneur.2014.00006> (2014).
2. von Hofsten, C. An action perspective on motor development. *Trends Cogn. Sci.* **8**, 266–272. <https://doi.org/10.1016/j.tics.2004.04.002> (2004).
3. Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J. & van der Wel, R. Cognition, action, and object manipulation. *Psychol. Bull.* **138**, 924–946. <https://doi.org/10.1037/a0027839> (2012).
4. Trevarthen, C. & Delafield-Butt, J. T. Autism as a developmental disorder in intentional movement and affective engagement. *Front. Integr. Neurosci.* **7**, 49. <https://doi.org/10.3389/fnint.2013.00049> (2013).
5. Hocking, D. R. & Caeyenberghs, K. What is the nature of motor impairments in autism, are they diagnostically useful, and what are the implications for intervention?. *Curr. Dev. Disord. Rep.* **4**, 19–27. <https://doi.org/10.1007/s40474-017-0109-y> (2017).
6. Anzulewicz, A., Sobota, K. & Delafield-Butt, J. T. Toward the autism motor signature: gesture patterns during smart tablet gameplay identify children with autism. *Sci. Rep.* **6**, 31107. <https://doi.org/10.1038/srep31107> (2016).
7. Crippa, A. *et al.* Use of machine learning to identify children with autism and their motor abnormalities. *J. Autism Dev. Disord.* **45**, 2146–2156. <https://doi.org/10.1007/s10803-015-2379-8> (2015).
8. Bone, D. *et al.* Applying machine learning to facilitate autism diagnostics: pitfalls and promises. *J. Autism Dev. Disord.* **45**, 1121–1136. <https://doi.org/10.1007/s10803-014-2268-6> (2014).
9. Rao, A., Monteiro, J. M. & Mourao-Miranda, J. Predictive modelling using neuroimaging data in the presence of confounds. *NeuroImage* **150**, 23–49. <https://doi.org/10.1016/j.neuroimage.2017.01.066> (2017).
10. Koul, A., Cavallo, A., Ansuini, C. & Becchio, C. Doing it your way: how individual movement styles affect action prediction. *PLoS ONE* **11**, e0165297. <https://doi.org/10.1371/journal.pone.0165297> (2016).
11. Ting, L. H. *et al.* Neuromechanical principles underlying movement modularity and their implications for rehabilitation. *Neuron* **86**, 38–54. <https://doi.org/10.1016/j.neuron.2015.02.042> (2015).
12. Richardson, M. J. & Johnston, L. Person recognition from dynamic events: the kinematic specification of individual identity in walking style. *J. Nonverbal Behav.* **29**, 25–44. <https://doi.org/10.1007/s10919-004-0888-9> (2005).
13. Patri, J.-F. *et al.* Transient disruption of the inferior parietal lobule impairs the ability to attribute intention to action. *Curr. Biol.* **30**, 4594–4605. <https://doi.org/10.1016/j.cub.2020.08.104> (2020).
14. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C. & Kording, K. P. The need to approximate the use-case in clinical machine learning. *Gigascience* **6**, 1–9. <https://doi.org/10.1093/gigascience/gix019> (2017).
15. Cavallo, A. *et al.* Prospective motor control obeys to idiosyncratic strategies in autism. *Sci. Rep.* **8**, 13717. <https://doi.org/10.1038/s41598-018-31479-2> (2018).
16. Chaibub Neto, E. *et al.* Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *Digit. Med.* **2**, 99. <https://doi.org/10.1038/s41746-019-0178-x> (2019).
17. Hernandez-Lemus, E., Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **14**, e0224365. <https://doi.org/10.1371/journal.pone.0224365> (2019).
18. Whelan, R. & Garavan, H. When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol. Psychiat.* **75**, 746–748. <https://doi.org/10.1016/j.biopsych.2013.05.014> (2014).
19. Jbabdi, S. *et al.* Multivariate classification of neuroimaging data with nested subclasses: biased accuracy and implications for hypothesis testing. *PLoS Comput. Biol.* **14**, e1006486. <https://doi.org/10.1371/journal.pcbi.1006486> (2018).
20. Cawley, G. C. & Talbot, N. L. C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **11**, 2079–2107. <https://doi.org/10.5555/1756006.1859921> (2010).
21. Torres, E. B. & Donnellan, A. M. Editorial for research topic “Autism: the movement perspective”. *Front. Integr. Neurosci.* **9**, 12. <https://doi.org/10.3389/fnint.2015.00012> (2015).
22. Casartelli, L., Molteni, M. & Ronconi, L. So close yet so far: motor anomalies impacting on social functioning in autism spectrum disorder. *Neurosci. Biobehav. Rev.* **63**, 98–105. <https://doi.org/10.1016/j.neubiorev.2016.02.001> (2016).
23. Cook, J. L., Blakemore, S.-J. & Press, C. Atypical basic movement kinematics in autism spectrum conditions. *Brain* **136**, 2816–2824. <https://doi.org/10.1093/brain/awt208> (2013).
24. Varlet, M., Marin, L., Lagarde, J. & Bardy, B. G. Social postural coordination. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 473–483. <https://doi.org/10.1037/a0020552> (2011).
25. Becchio, C., Sartori, L. & Castiello, U. Toward you: the social side of actions. *Curr. Dir. Psychol. Sci.* **19**, 183–188 (2010).
26. Trujillo, J. P., Simanova, I., Bekkering, H. & Özyürek, A. Communicative intent modulates production and comprehension of actions and gestures: a Kinect study. *Cognition* **180**, 38–51. <https://doi.org/10.1016/j.cognition.2018.04.003> (2018).
27. Sinha, P. *et al.* Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. U S A* **111**, 15220–15225. <https://doi.org/10.1073/pnas.1416797111> (2014).
28. Mason, A. H. & MacKenzie, C. L. Grip forces when passing an object to a partner. *Exp. Brain Res.* **163**, 173–187. <https://doi.org/10.1007/s00221-004-2157-x> (2005).
29. Bates, C. J., Yildirim, I., Tenenbaum, J. B. & Battaglia, P. Modeling human intuitions about liquid flow with particle-based simulation. *PLoS Comput. Biol.* **15**, e1007210. <https://doi.org/10.1371/journal.pcbi.1007210> (2019).
30. Arthur, T., Vine, S., Brosnan, M. & Buckingham, G. Predictive sensorimotor control in autism. *Brain* **143**, 3151–3163. <https://doi.org/10.1093/brain/awaa243> (2020).
31. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20**, 365–377. <https://doi.org/10.1038/nn.4478> (2017).
32. Wechsler, D. *Wechsler Intelligence Scale for Children*, 4th edn (Harcourt Assessment, San Antonio, TX, 2003).
33. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn (American Psychiatric Publishing, Arlington, VA, 2013).
34. Lord, C. *et al.* *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1–4* (Western Psychological Services, Torrance, CA, 2012).
35. Rutter, M., Le Couteur, A. & Lord, C. *The Autism Diagnostic Interview-Revised (ADI-R)*. (Western Psychological Services, Torrance, CA, 2003).
36. Anderson, P., Anderson, V. & Lajoie, G. The tower of London test: validation and standardization for pediatric populations. *Clin. Neuropsychol.* **10**, 54–65. <https://doi.org/10.1080/13854049608406663> (1996).
37. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4) (1971).
38. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* **310**, 2191–2194. <https://doi.org/10.1001/jama.2013.281053> (2013).

Acknowledgements

This research was supported by the European Research Council under the ERC-2017-PoC Grant Agreement 789601. DINO GMI contributed to this work within the framework of the DINO GI Department of Excellence MIUR 2018-2022 (legge 232 del 2016). We thank all participants and their families for their efforts to participate in the study. We also thank Marco Jacono and Jessica Podda for their help with data collection and preprocessing, and Michael Lombardo for useful comments.

Author contributions

C.B. and A.C. conceived of the study. F.B. recruited patients with L.N. C.A. and F.B. tested the participants and collected the data. S.P., A.C., L.R. and C.B. designed the analyses. A.C., L.R. and S.P. analysed the data. A.C., S.P. and C.B. interpreted results. M.P. advised on aspects of machine learning. C.B., A.C. and S.P. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-82374-2>.

Correspondence and requests for materials should be addressed to C.B.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021