Attributing Minds to Triangles: Kinematics and Observer-Animator Kinematic Similarity predict Mental State Attribution in the Animations Task

Bianca A. Schuster[1*], Dagmar S. Fraser[1], Jasper J.F. van den Bosch[1], Sophie Sowden[1], Andrew S. Gordon[2], Dongsung Huh[3], and Jennifer L. Cook[1]

[1] University of Birmingham, School of Psychology, Birmingham, B15 2TT, United Kingdom

[2] University of Southern California, Institute for Creative Technologies, Los Angeles, CA 90094, United States

[3] MIT-IBM Watson AI Lab, Cambridge, MA 02142, United States

Author Notes

Bianca Schuster: Orchid ID https://orcid.org/0000-0001-5584-2078

Correspondence concerning this article should be directed to Bianca Schuster

biancaschuster05@gmail.com

Acknowledgements

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under ERC-2017-STG Grant Agreement No 757583 (Brain2Bee; Jennifer Cook PI).

We thank Bodo Winter, Hélio Cuve, Jo Cutler and Olivier Codol for their assistance with statistical questions.

ABSTRACT

The ability to ascribe mental states, such as beliefs or desires to oneself and other individuals forms an integral part of everyday social interaction. One task that has been extensively used to test mental state attribution in a variety of clinical populations is the animations task, where participants are asked to infer mental states from short videos of interacting triangles. In this task, individuals with clinical conditions such as autism spectrum disorders typically offer fewer and less appropriate mental state descriptions than controls, however little is currently known about why they show these difficulties. Previous studies have hinted at the similarity between an observer's and the triangles' movements as a key factor for the successful interpretation of these animations. In this study we present a novel adaptation of the animations task, suitable to track and compare animation generator and -observer kinematics. Using this task and a population-derived stimulus database, we demonstrate that an animation's kinematics and kinematic similarity between observer and generator are integral for the correct identification of that animation. Our results shed light on why some clinical populations show difficulties in this task and highlight the role of participants' own movement and specific perceptual properties of the stimuli.

Introduction

Seminal work by Heider and Simmel[1] demonstrated that humans readily attribute mental states to two triangles moving around a rectangular enclosure. Since their inception in 1944 such “animations tasks” (also referred to as Frith-Happé Animations[2] and Social Attribution Task[3]) have grown dramatically in popularity and have been used in a wide variety of clinical populations, including autism spectrum disorder (ASD)[2,4], Schizophrenia[5], antisocial personality disorder[6], Huntington’s disease[7] and Tourette’s syndrome[8]. Though animations tasks have been scored and administered in a number of ways (Some studies count the number of mental state terms used to describe the movements of the triangles[2,4], other studies have asked participants to rate the type of interaction or the mental state word depicted in the animations[9,10]) it is generally agreed that “poor performance” indicates a problem with identifying the triangles as mentalistic agents and ascribing appropriate mental states to them. We refer to these processes here as ‘mental state attribution’.

Though mental state attribution has been found to be atypical across a range of clinical populations, little is known about *why* some individuals struggle to attribute appropriate mental states to the triangles. One explanation is that individuals who struggle with the animations task would exhibit atypicalities in other tests of mental state attribution because of a deficit in the ability to attribute minds and ascribe appropriate mental states. However, animations tasks tend to be more sensitive to mental state attribution difficulties compared to other tests, as shown by Abell et al.[2].

A recent study highlights that kinematic similarities between the triangles’ movements and the participant’s own movements may influence performance on the animations task[9]. Edey and colleagues asked autistic (‘condition-first’ terminology is used in line with the majority preference expressed in a survey of the autistic community[11]) and non-autistic participants to complete the animations task, and also to produce their own animations using triangles that could be moved around an enclosure with magnetic levers. The authors found that animations produced by autistic individuals were more jerky (i.e. exhibited greater changes in acceleration and deceleration) than those produced by non-autistic individuals. Furthermore, whereas non-autistic participants could readily attribute mental states to animations created by other non-autistic participants, they had difficulties attributing mental states to the jerky animations that had been produced by the autistic participants. The authors proposed that *movement similarity* significantly contributes to performance in the animations task: that is, non-autistic individuals were better able to correctly identify animations created by other non-autistic participants because the movement kinematics in the videos were similar to the kinematics that they themselves would use to move the triangles. Conversely, autistic participants in in Edey's study did not show improved performance when rating their own group's relative to the control group's animations. The authors concluded that the increased variability in jerk present within this group lead to a reduced number of animations sufficiently similar to facilitate mentalizing performance in their autistic participants.

The proposal that movement similarity may affect performance in the animations task is bolstered by recent empirical work showing that observers more accurately estimate a human actor's underlying intentions when the kinematics of the actor's movements closely approximate the observer's own movement kinematics[12]. Furthermore, a role for movement similarity in mental state attribution is in line with theoretical accounts suggesting that inferences about others' actions are facilitated by mapping visual representations of others' actions onto our own visual/motoric representations of the same actions[13-16]. The movement similarity hypothesis would propose that mental state attribution difficulties in classic animations tasks may, at least in part, be explained by differences between the way the triangles are animated and the way an observer would move the triangles if required to create their own animation. This raises the possibility that clinical groups might exhibit accurate mental state attribution for animations where kinematics are matched to a participant's own movement kinematics. To better understand why some individuals struggle to attribute appropriate mental states in the animations task, the first aim of the current study was to test the hypothesis that a significant amount of variance in performance in the animations task would be accounted for by the kinematic jerkiness of the animation and the *similarity* between the kinematics of the animation and a participant's own movements.

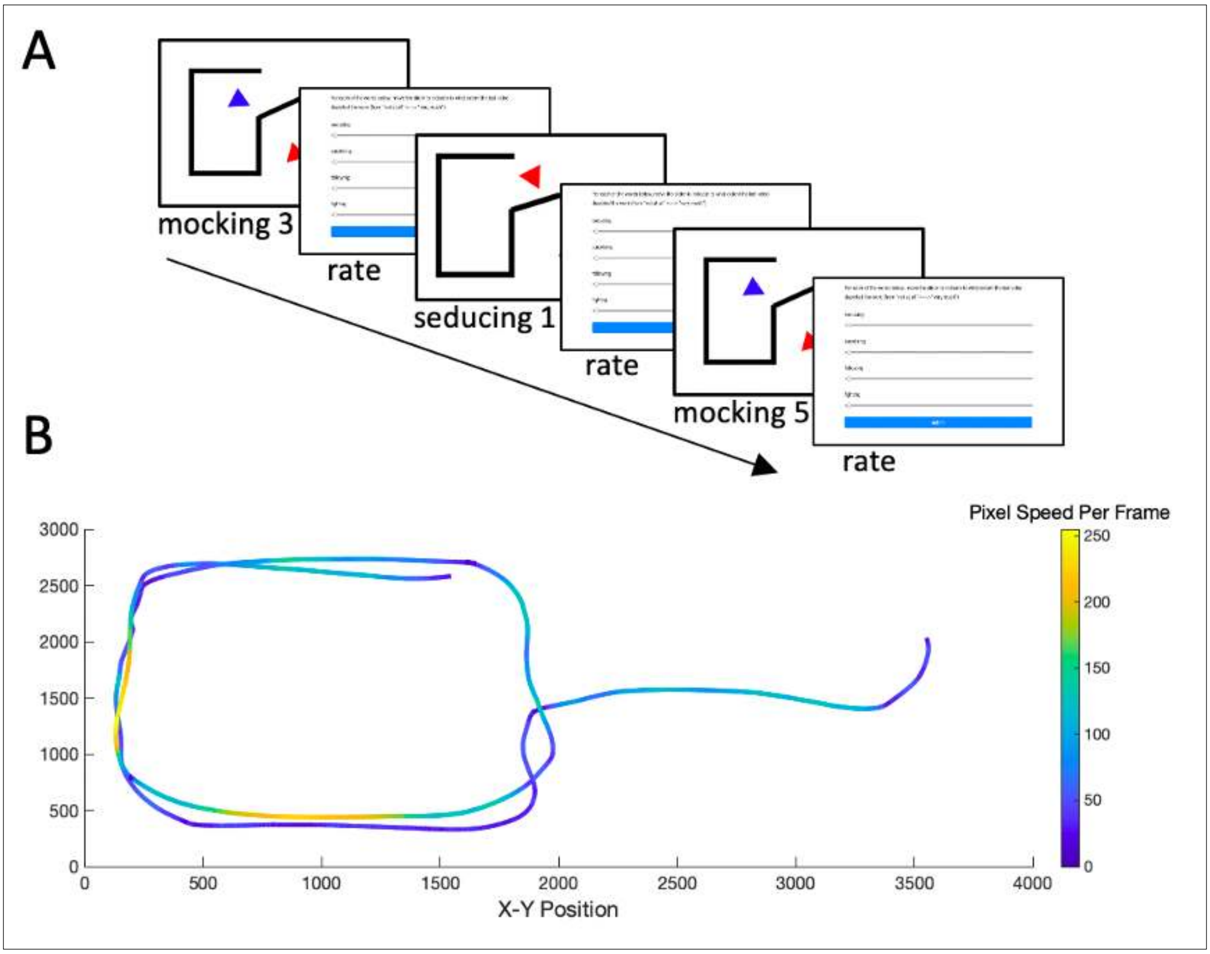
Kinematic jerk and movement similarity are not the only factors which plausibly influence performance on the animations task. Previous studies have highlighted potential roles for stimulus features including the rotation of, and distance between, the triangles[17], and the shape of the triangles' trajectories[18]. For instance, Roux et al. documented highly distinguishable trajectory paths for random, goal-directed and mental state animations, thus suggesting that trajectory path may be an important cue in mental state attribution. Correspondingly, the second aim of the current study was to explore the extent to which a range of other stimulus features, including trajectory shape, influence the ease with which participants correctly attribute a mental state to an animation. By doing so, we shed light on a multiplicity of factors which may explain why some clinical groups find the animations task so challenging.

For this latter analysis we made use of the fact that, similar to a sound wave, a triangle's trajectory comprises a complex wave and thus can be decomposed with Fourier transform and represented as spectral density in different frequency bands[19]. In other words, Fourier transform can be used to characterize the shape of a trajectory. For example, a trajectory which approximately follows an elliptical orbit oscillates in speed and curvature twice during every full rotation and consequently would be characterized by high spectral density in a band centered around an angular frequency of two. Adapting a method developed by Huh & Sejnowski we explored whether there are particular angular frequency bands which differentiate mocking, seducing, surprising, following and fighting animations and whether spectral density in these bands was predictive of accuracy.

Currently available animation task datasets are not suitable to test our hypotheses for two reasons: First, having been created by experimenters or graphic designers, the stimuli in these tasks typically represent a narrow range of kinematics and thus lack the variation necessary for quantifying the contribution of kinematics and other stimulus features to performance. Second, tasks to date offer no option to track animator (or observer) kinematics at sufficient sampling rates to reliably make inferences about the role of movement similarity. Here we created a novel animations database (available upon request) by asking 51 members of the general population to animate two triangles to depict mental- (mocking, seducing, surprising) and non-mental- (following, fighting) state interactions on a 133 Hz touch screen device. Subsequently an independent sample of 37 members of the general population watched a selection of videos from our new database. To ensure that participants were exposed to a wide range of kinematics they watched 8 exemplars, for each word, ranging from slow to fast speed. Participants rated the extent to which each animation depicted the words mocking, seducing, surprising, following and fighting, in addition to also creating their own animation for each word (Fig. 1). In a three-step analysis procedure, we first used Bayesian mixed effects models to test our hypotheses that kinematic jerk and the similarity in kinematics between observer and animator are significant predictors of the accuracy of mental state attributions (*confirmatory analysis*). In a second step, we used Fast Fourier Transform (FFT) combined with bootstrapped F-tests to investigate whether mocking, seducing, surprising, following and fighting animations could be reliably distinguished according to the profile of spectral density across a range of frequency bands (*exploratory analysis 1*). Finally, we employed random forest analysis to determine the relative contribution to accuracy of a multiplicity of factors including speed, acceleration, jerk, the amount of simultaneous movement of both triangles, the relative distance between triangles, triangles' average rotation and the magnitude of *spectral density* in the frequency bands identified in the second analysis step (*exploratory analysis 2)*.

Figure 1

(A) Schematic depiction of three successive trials in the animations task. (B) Example trajectory of an animation stimulus.



Note. (A) 37 participants watched videos from the database and rated the extent to which each video depicted mocking, seducing, surprising, following, or fighting. (B) Each participant used a touchscreen device to create their own triangles animations. For each animation (both observed and generated by participants) we calculated *jerk* as the mean of the third order non-null derivative of the raw positional data across all frames, movement similarity was calculated as the difference in mean jerk between an animation stimulus and the participant's own animation of the same word (*jerk difference*). Depicted is an example of a *following* animation (one triangle's trajectory).

Results

Accuracy for each trial was calculated by subtracting the mean rating for all non-target words from the rating for the target word (e.g., the target word was *seducing* on trials where the participant watched a video wherein the original animator had attempted to depict the triangles seducing each other). Consequently, a high, positive accuracy score for a seducing animation indicates that an observer rated this animation as depicting seducing to a higher extent than mocking, surprising, following or fighting. For a comparison of mean accuracy scores for each word category see Supplementary Materials. For each video that participants observed and for each animation that they created themselves, mean jerk magnitude (hereafter: *jerk*) was obtained by taking the third order non-null derivatives of the raw positional data and calculating the mean across all frames in the video. Movement similarity was calculated as the difference in mean jerk between an animation stimulus and the participant's own animation of the same word (hereafter: *jerk difference*), where lower difference values indicate higher movement similarity (see **Methods: Data Analysis and Processing**).

Mental state animations are rated less accurately than non-mental state animations

The distinction between mental state and non-mental state, and the individual words to depict these two conditions, are equivalent to the Theory of Mind and Goal-Directed conditions used in the original paradigm by Abell et al.[2], and have since been widely used across the literature[4,9,10]. A Bayesian linear mixed effects model with the maximal random effects structure allowed by the design[20] (random intercepts for *animation ID* (unique identifier for each animation) and *subject ID;* random slopes for all fixed effects varying by *animation ID* and *subject ID*) was fitted to jerk, jerk difference (lower values reflect higher movement similarity) and the dummy-coded factor *mental state* (mental state [seducing, surprising, mocking] versus non-mental state [following, fighting]) as well as their three-way interaction.

For all relevant model parameters, we report expected values (E_μ) under the posterior distribution and their 95% credible intervals (CrIs)[21], as well as the posterior probability that an effect is different to zero ($P(E_\mu < 0)$ / $P(E_\mu > 0)$). In line with Franke & Roettger[22], if a hypothesis states that an effect $E_\mu \neq 0$ (e.g. effect of movement similarity on accuracy), we conclude there is compelling evidence for this effect if zero is not included in the 95% CrI of E_μ and if the posterior probability P($E_\mu \neq 0$) is close to 1.

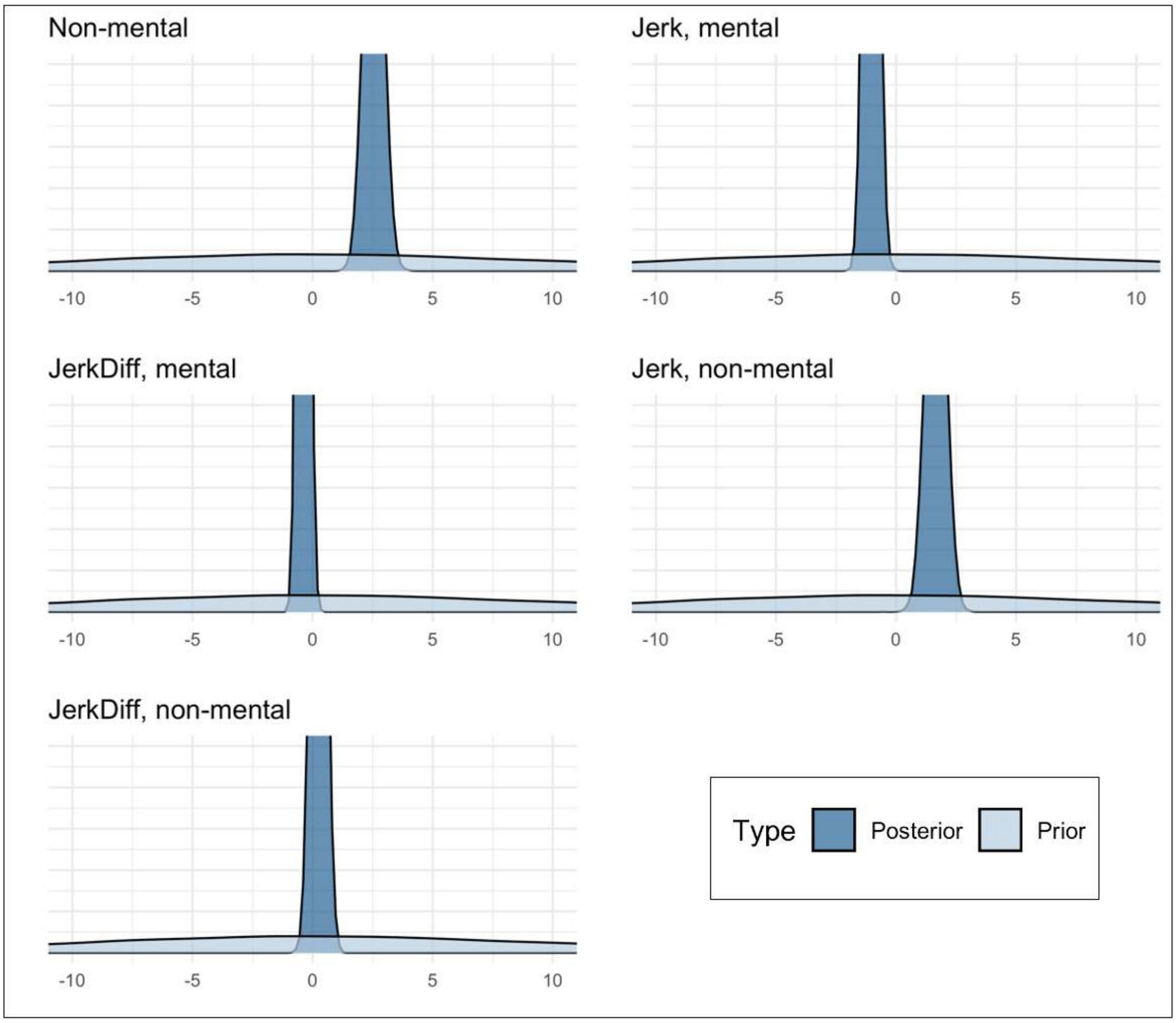
The model indicated that accuracy was higher in non-mental state videos relative to mental state videos ($E\mu_{non-mental}$ = 2.54, CrI= [1.81, 3.28]), with the posterior probability that the effect is larger than zero being P($E\mu_{non-mental} > 0$) = 1 (see Fig. 2 for prior and posterior distributions of all estimated parameters).

Jerk affects performance differently for mental- and non-mental state animations

In line with our hypothesis, accuracy was associated with mean jerk, furthermore jerk interacted with mental state: For mental state animations, lower mean jerk was associated with higher accuracy ($E\mu_{jerk,mental}$ = -1.03, CrI = [-1.52, -0.53]), whereas in non-mental state animations higher mean jerk led to higher accuracy scores ($E\mu_{jerk,non-mental}$ = 1.65, CrI = [0.88, 2.41]). Thus, while mental state animations with mean jerk values higher than 1 standard deviation (SD) above the mean were rated 1.03 points less accurately, in non-mental state animations higher jerk values increased accuracy by 1.65 points. Since the posterior probabilities for both effects (P($E\mu_{jerk,non-mental} > 0$), P($E\mu_{jerk,mental} < 0$)) were in fact 1, we conclude that, given our model and the data, there is compelling evidence in favor of our hypothesis that an animations' jerk impacts mental state attribution performance in the animations task. To probe whether such effects varied as a function of the word depicted in the video, we conducted separate exploratory models for non-mental state and mental state animations for which we included *word category* (non-mental state: following, fighting; mental

Figure 2

Prior and posterior probabilities of model parameters predicting accuracy



Note. JerkDiff = *jerk difference*. For all regression coefficients, weakly informative priors were set as following a normal distribution centered at 0 with an SD of 10.

state: mocking, seducing, surprising) as a predictor in addition to jerk and jerk difference. These models revealed that, for non-mental state animations there was a strong negative effect of jerk for fighting, but not following, animations ($E\mu_{jerk,fighting}$ = 1.88, CrI = [0.67, 3.11], P($E\mu_{jerk,fighting} > 0$) = 1; $E\mu_{jerk,following}$ = 0.30, CrI = [-0.30, 1.05]). For mental state animations, the overall negative effect of jerk was driven by a tendency towards a negative effect of jerk on accuracy in mocking and surprising animations ($E\mu_{jerk,mocking}$ =

-0.58, CrI = [-1.56, 0.40]; $E\mu_{jerk,surprising}$ = -0.94, CrI = [-2.69, 0.76]). There was no effect of jerk in seducing animations ($E\mu_{jerk,seducing}$ = 0.26, CrI = [-1.40, 1.85]).

Higher observer-animator similarity in jerk is associated with higher accuracy only in mental-state animations

In line with our hypothesis, accuracy was also associated with jerk difference, furthermore jerk difference interacted with mental state such that it was a significant predictor for mental, but not non-mental, state videos. That is, for non-mental state animations the mean of all posterior coefficients for jerk difference was centered near zero ($E\mu_{jerkDiff,non-mental}$ = 0.25, CrI = [-0.27, 0.76]). In contrast, for mental state animations the credible interval of jerk difference did not include zero ($E\mu_{jerkDiff,mental}$ = -0.38, CrI = [-0.72, -0.03]) and the estimated probability of this effect being below zero (P ($E\mu_{jerkDiff,mental} < 0$)) was 0.98. Thus, jerk difference had a negative effect on accuracy for mental state animations only. Consequently, in mental state animations, *higher* movement similarity was associated with higher accuracy. To probe whether such effects varied as a function of word category we conducted an exploratory mixed model which included the word categories mocking, seducing and surprising. This model revealed that jerk difference affected performance only in mocking animations ($E\mu_{jerkDiff,mocking}$= -0.70, CrI = [-1.22, -0.18]; P ($E\mu_{jerkDiff,mocking} < 0$) = 0.99; $E\mu_{jerkDiff,seducing}$= 0.98, CrI = [-0.49, 2.46]; $E\mu_{jerkDiff,surprising}$= 0.63, CrI = [-0.29, 1.52]).

A combination of ten kinematic and spatial variables best predicts accuracy in the animations task

To investigate whether different triangle trajectories can reliably distinguish between the five target words (i.e., mocking, seducing, surprising, following, fighting) we used FFT to decompose the triangles' trajectories and represent them as an amplitude spectral density profile across a range of angular frequencies. To test for differences, between the five target words, in spectral density across the angular frequency range, bootstrapped F-tests (with 1000 boots) were performed (see **Methods: Data Analysis and Processing**). This analysis revealed nine significant clusters, defined as clusters of difference that occurred in less than 5% of comparisons with resampled distributions (see Figure 3A).

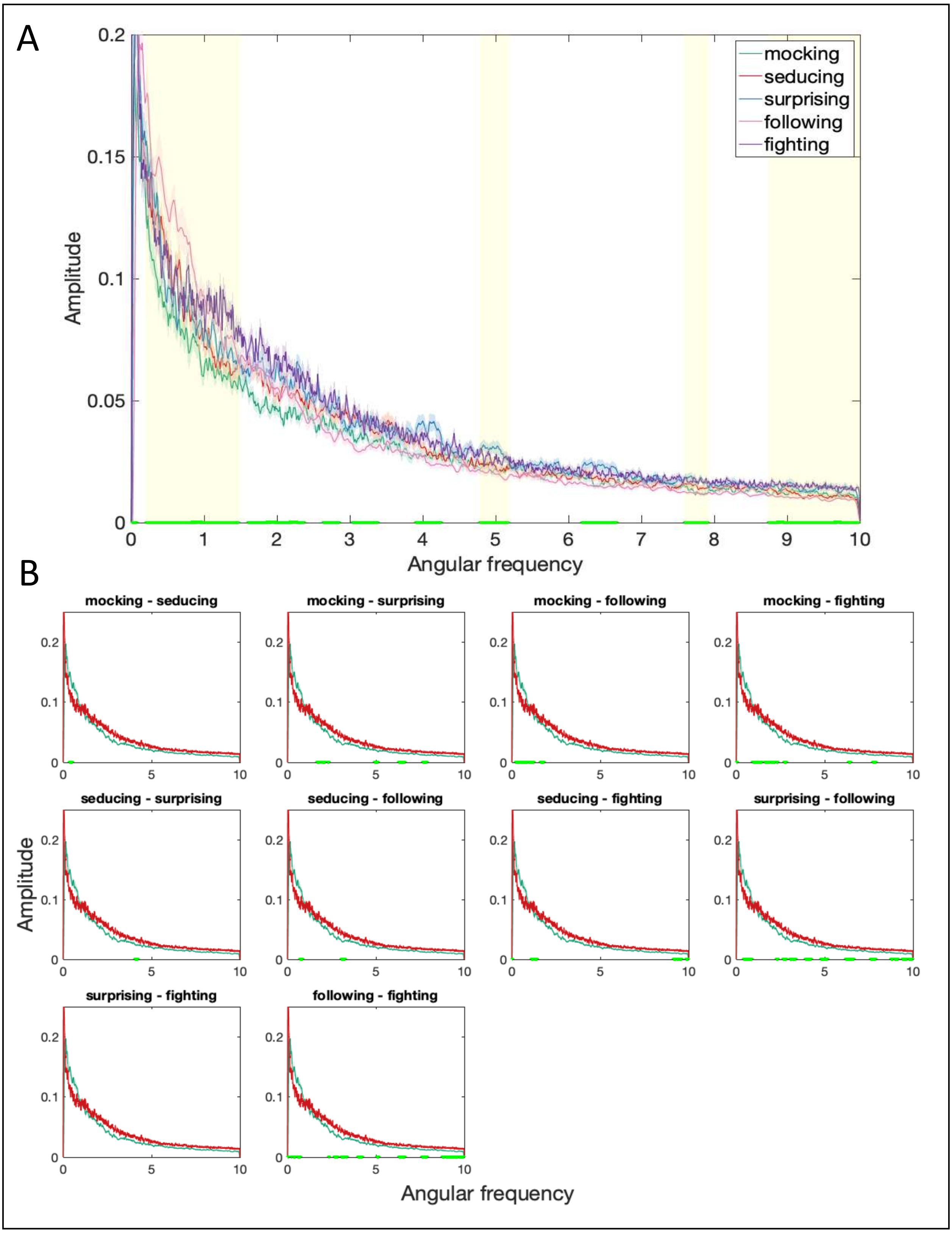
To examine whether spectral density in these nine frequency clusters was predictive of accuracy we used the maxima and minima of each significant cluster as bin edges and calculated the *angular frequency spectral density (AFSD)* as the area under the curve between the bin edges (cluster bin edges: 0.21 – 1.49, 1.61 – 2.39, 2.64 – 2.87, 3.04 – 3.40, 3.91 – 4.27, 4.79-5.19, 6.19-6.68, 7.6-7.93, 8.75-10). The relative contribution to accuracy of AFSD in bins 1-9 was assessed, alongside mental-state, speed, acceleration magnitude (hereafter: *acceleration*), jerk, *simultaneous movement*, *relative distance* and *mean rotation*, by means of a random forest model[23] using the *Boruta*[24] wrapper algorithm (version 7.7.0). Boruta trains a random forest regression model on all variables as well as their permuted copies - so called "shadow features" - and classes a variable as *important* when its permutation importance is significantly higher than the highest permutation importance of a shadow feature (for more details see **Methods: Exploratory analysis**). Note that because this analysis technique does not account for random effects, values corresponding to the same animation were averaged across participants, this permits indices such as jerk and acceleration which are features of a particular animation but excludes jerk difference which depends on the relation between an animation and an individual participant.

Out of all 16 variables tested, 10 were confirmed *important*, two were confirmed *unimportant*, and four were classed as *tentative* on the basis that their permutation importance was not significantly different from the maximal importance of a shadow feature (see Fig 4). Fig 4 illustrates that the important role of mental-state and jerk in predicting accuracy is confirmed by the random forest model, with mean importances of 16.0 and 7.82 respectively. However, the model identifies a third variable as even more important than jerk: mean rotation (mean importance = 11.78). In addition, an animation's acceleration and speed, AFSD in bins 1, 6, 9 and 8, as well as the amount of simultaneous movement of both triangles notably contribute to explaining performance in the animations task (mean importances: acceleration = 7.91; speed = 4.70; AFSD-bin 1 = 7.03, AFSD-bin 6 = 6.37, AFSD-bin 9 = 5.04, AFSD-bin 8 = 3.89; simultaneous movement = 4.74). A final model of all 10 important variables predicting accuracy was evaluated by training a random forest on a subset of 70% of the data (training set) and using it to predict the remaining 30% (test set). The regression model of the training set predicting the test set was highly significant ($p < .001$) and indicated that the selected variables explained 37% of accuracy values.

We subsequently conducted post hoc random forests separately for mental state- and non-mental state animations. These post hoc analyses revealed that, in mental state animations, five factors were predictive of accuracy, with jerk and acceleration being the most prominent predictors, followed by speed, which was ranked third (see Supplementary Fig 2). In addition, AFSD in bin 6 and simultaneous movement were classed as important in predicting accuracy. In non-mental state animations, a total of eight predictors were identified as important variables, with mean rotation being ranked highest by a considerable margin. In addition to mean rotation, a combination of AFSD in bins 1, 6, 7 and 9, and acceleration, jerk and speed were identified as important features of non-mental state animations.

Figure 3

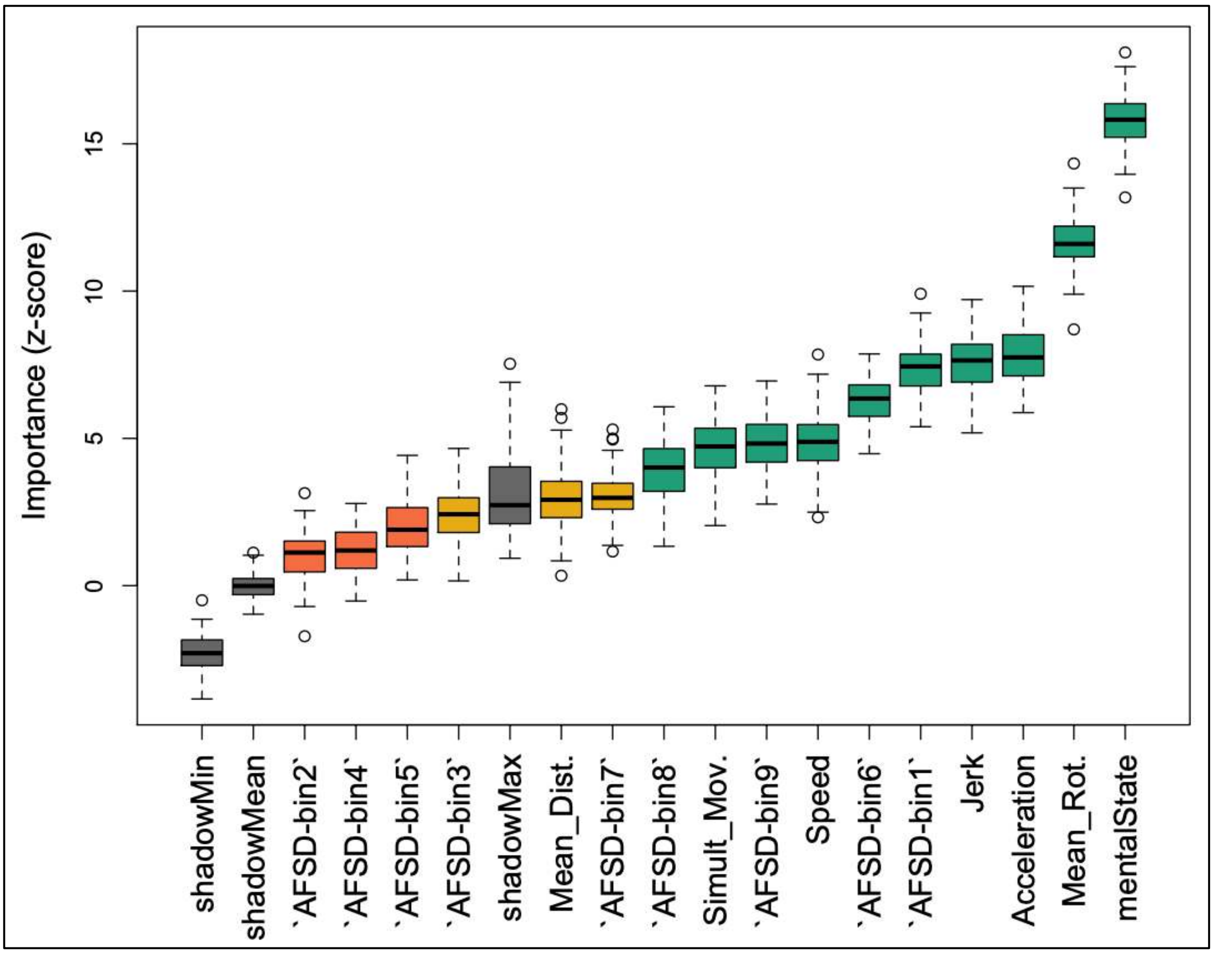
Significant clusters of difference in angular frequency spectral density (AFSD)



Note. A) Solid colored lines represent spectral density as a function of angular frequency per word (=AFSD), the corresponding shaded areas represent 1 SEM (standard error of the mean) below and above the mean values. Yellow bars on x-axis represent clusters where AFSD significantly differs between mocking, seducing, surprising, following and fighting. Clusters that are predictive of accuracy are highlighted in yellow. Note that the lowest angular frequency derived from the data varied between 0.02 and 0.09, resulting in extrapolated values for some participants. For this reason, the first cluster of difference ranging from 0.02 to 0.09 was considered not representative of actual movements and disregarded. B) Post-hoc comparisons of AFSD.

Figure 4

Random forest variable importances



Note. Variable importances of all 16 features entered into the Boruta random forest, displayed as boxplots. Box edges denote the interquartile range (IQR) between first and third quartile; whiskers denote 1.5 * IQR distance from box edges; circles represent outliers outside of 1.5 * IQR above and below box edges. Box color denotes decision: Green = confirmed, yellow = tentative, red = rejected; grey = meta-attributes shadowMin, shadowMax and shadowMean (minimum, maximum and mean variable importance attained by a shadow feature).

Discussion

To better understand why some clinical groups find the animations task so challenging, this study evaluated the relative contribution of jerk, jerk similarity and other stimulus characteristics to mental state attribution performance. Our results confirm our hypothesis that kinematic jerk and movement similarity are predictors of the accuracy of mental state attribution. In addition, we highlight that stimulus features including the shape of the triangles’ trajectories and the amount of rotation of the triangles can also affect the ease with which participants are able to appropriately label the target states.

In the first part of our three-step analysis, we found that mental state was the primary predictor of animations task performance. Mental state videos were strongly associated with lower accuracy, correspondingly non-metal state videos were rated more accurately. The observation that our healthy participants performed worse when interpreting mental state animations is inconsistent with previous findings. In Abell et al.'s and other studies, non-autistic adult participants performed equally well on non-mental state and mental state animations[2,4,25]. It is possible that our findings illustrate a true difference in difficulty between mental and non-mental state attribution that is revealed only when participants are presented with a wide range of animation stimuli from a population-derived database. This difference may have been overlooked because previous studies employed animations created by a single graphic designer, or small group of experimenters and thus lack variation. However, this possibility demands empirical testing. Indeed, a direct comparison between our paradigm and previous studies is not possible due to task related differences (e.g., in indices of performance, and number of words animated per condition).

In this first analysis step it was further revealed that the triangles' mean jerk in an animation plays a substantial role in interpreting that animation. For mental state attributions jerk was *negatively* predictive of accuracy, whereas for non-mental state animations jerk was *positively* predictive of accuracy. Post hoc analyses revealed that this latter result was primarily driven by fighting animations, and that the former was most notable with respect to mocking and surprising animations (though caution is advised since credible intervals of coefficient estimates did not exclude zero). In previous work, Edey and colleagues[9] observed that non-autistic participants were more accurate in their mental state attributions for animations generated by non-autistic participants compared to those generated by autistic participants.

They also observed that animations generated by autistic participants were more jerky compared to those generated by controls. However, in Edey et al.'s study there were a number of additional dimensions along which the two groups' animations may have varied, making it impossible to know whether the autistic participants' animations were difficult to interpret *because of* the jerky kinematics. Our results show that jerk meaningfully contributes to the accuracy of mental state attributions, thus our data supports the conclusion that jerk is highly likely to be one of the driving factors in the group differences observed by Edey et al.

Our results also highlight kinematic similarity as a potential driving factor of the differences observed by Edey et al.[9]. That is, we observed a positive relationship between kinematic similarity and accuracy. Post hoc analyses revealed that evidence of this relationship was particularly compelling in the case of mocking animations: The more closely a mocking animation's mean jerk approximated the participant's own jerk when animating the same word category, the more accurately that animation was rated. We speculate that Edey et al.'s non-autistic participants performed poorly when attributing mental states to animations produced by autistic individuals not only because these animations were jerky, but also because the kinematics of the animations were *dissimilar* from the way in which the observer would have produced the same animation.

The second aim of the current study was to explore the extent to which a range of other stimulus features, including trajectory shape, influence mental state attribution accuracy. To quantify trajectory shape we used FFT to decompose trajectories into spectral density in angular frequency bins. Animation identity could be differentiated by AFSD in nine bins and random forest analyses confirmed that four of these bins - bins 1, 6, 8 and 9 corresponding to angular frequencies 0.2-1.5, 4.8-5.2, 7.6-7.9, 8.8-10 - were 'important' predictors of mental state attribution accuracy. Relative to the other words, following animations had the highest AFSD in the angular frequency range 0.2-1.5 (bin 1; Fig. 3). A high amount of AFSD in this range indicates a trajectory characterized by complex doodle-like movements (See Supplementary Fig. 3) with low angular-frequency oscillation in speed and curvature. Thus, one may speculate that animations which are most easily identifiable as ‘following’ comprise doodle-like triangle trajectories, with between 0.2 and 1.5 curvature oscillations per 2π radians. In the angular frequency range 4.8-5.2 (bin 6), surprising animations had highest AFSD relative to the other words (See Fig. 3). This angular frequency range centers around the pure-frequency trajectory of a pentagon and thus is reflective of movements with around five speed-curvature oscillations per 2π radians. Whilst our stimuli did not necessarily contain trajectories in the shape of actual pentagons, high AFSD in bin 6 reflects curves and speed-curvature patterns similar to those required to produce a closed-form pentagon. Finally, relative to the other words, both surprising and fighting had high AFSD in angular frequency ranges 7.6-7.9 (bin 8) and 8.8-10 (bin 9). A high amount of AFSD in these ranges indicates trajectories characterized by octagonal (bin 8) and decagonal shapes (See Fig. 4) with 8-10 speed-curvature oscillations per rotation. Together these results clearly illustrate that trajectory shape comprises an important cue with respect to the identity of the word that is depicted in an animation. At present one can only speculate about why some shapes (e.g., pentagons) are more indicative of particular mental/non-mental states (e.g., surprising).

For the third step in our three-part analysis, we employed random forests to ascertain the relative contribution to accuracy of a range of stimulus features. The random forest methodology was chosen for its robustness against (multi-)collinearity and suitability for evaluating contributions of a large number of variables with limited data points[26]. Our random forest analysis confirmed ten features as important predictors of accuracy. In order of relative importance these are: mental state, mean rotation, acceleration, jerk, trajectory shape (AFSD in bins 1, 6, 8, 9), simultaneous movement of the triangles and speed. Post hoc analyses (see Fig 3B) revealed that with respect to mental state attribution specifically, five of these features were of confirmed importance: jerk, acceleration, speed, AFSD-bin 6 and simultaneous movement. There was one feature which was uniquely important for mental state accuracy: The amount of simultaneous movement of blue and red triangles. By decomposing the animations task into features which predict accuracy, this random forest analysis deepens understanding of individual differences in animations task performance and raises testable empirical hypotheses for further research. For example, our analysis illustrates that simultaneous movement of the triangles is a stimulus feature which predicts mental state attribution accuracy. This observation raises the possibility that poor performance on the animations task in some clinical groups may be related to differences in processing this stimulus feature. That is, processing the simultaneous movement of the triangles requires distributed attention to two objects simultaneously. It may be that individuals with some clinical conditions exhibit a deficit in the perception of global relative to local motion stimuli (e.g., autism[27]) making it more difficult for them to process the simultaneous movement of two triangles. Here we show that this perceptual processing style would impact selectively on the accuracy of mental-, not non-mental-, state attributions.

Furthermore, our random forest analysis also raises interesting questions for further study. Since the random forest technique does not account for random effects, values corresponding to the same animation had to be averaged across participants, meaning that only features of a particular animation (e.g., jerk, speed) could be included and indices such as movement similarity, which depend on the *relation between* an animation and an individual participant were excluded. Future experiments are therefore required to investigate whether, similar to the jerk similarity effect we observed, there are also ‘similarity effects’ with respect to features such as simultaneous movement and trajectory shape. One may hypothesize that participants should be better able to infer mental states from animations which follow trajectories that are similar to the shapes they would produce themselves. Such an analysis has the potential to provide a clearer mechanistic understanding of atypical animations task performance in clinical groups. For example, given differences in upper limb- and fine motor control[28-31] autistic people may produce different trajectory shapes when creating their own animations. It remains to be seen whether apparent mentalizing deficits in autism are ameliorated when autistic people are provided with stimuli which match closely to features of their own movement including trajectory shape as well as kinematics.

The present findings highlight particular kinematic- and trajectory features as being critical for mental state attribution in the context of the animations task. This raises the possibility that individual differences in mentalizing may be related to individual differences in the perceptual processing of kinematics and trajectory information. Our findings further show that kinematic similarity between observer and animator facilitates mental state attribution. Consequently, individuals with certain clinical conditions might find the animations task particularly difficult due to differences in perceptual processing and/or reduced movement similarity. Our data paves the way for studies which empirically test whether mentalizing deficits in clinical populations persist when participants are provided with stimuli which closely match features (including kinematics, trajectory shape and amount of simultaneous movement) of their own movements.

Methods

Building the animotions database

Animotion Online Task

We created a browser-based application that enables us to record and replay participants' animations in the style of Heider & Simmel's original movies[1] while capturing the triangles' positions at 133Hz. For this purpose, we adapted a web application developed by Gordon & Roemmele (*The Heider-Simmel Interactive Theatre*[32], https://hsit.ict.usc.edu/) to fit our task design and allow instantaneous calculation of mean speed, acceleration and jerk (change in acceleration), thus enabling the selection of stimuli according to predefined criteria for replay. Gordon's web application employs scalable vector graphics (SVG) objects that allow simultaneous translation and rotation of each object with input from a single finger per object. To ensure object motion follows the direction of movement of the finger, and to suppress sporadic rotations (which can occur if dragging is initiated too close to the object center), object motion is suppressed until the pointer is dragged sufficiently far away from the center point (see https://asgordon.github.io/rotodrag-js/ for a more detailed description of the library used for this application).

Participants

We asked 51 healthy volunteers (46 females, mean (M) [SD] age = 20.23 [2.71] years, range 18-34 years) to animate two triangles in order to depict three mental state (mocking, seducing, surprising) and two non-mental state (following, fighting) words. Participants were recruited from the University of Birmingham research participation scheme, gave written informed consent and received either course credit or money (£8 per hour) for their participation. All experimental procedures were conducted in line with the WMA declaration of Helsinki[33] and approved by the University of Birmingham Research Ethics Committee (ERN 16-0281AP5).

Procedure

Data was collected at the University of Birmingham. Individuals were seated in front of a WACOM Cintiq 22 HD touch screen, tilted at an angle of approximately 30 degrees upon the desk. They were presented with the starting frame, comprising a black rectangular enclosure and two equally sized equilateral triangles (one red and one blue) on a white background (see Supplementary Figure 4). In a 45-second-long practice phase, participants familiarized themselves with how to use their finger movements in order to navigate the triangles around the screen. Participants were subsequently instructed to '*represent certain words by moving the triangles around the screen*', assured they could move the triangles in any way they saw fit and encouraged to use their index fingers on both the left and right hand to move the triangles simultaneously (for a complete transcript of task instructions see Supplementary Materials). A dictionary was provided in case participants did not know the word in question. No further explanations were given.

Following instructions, participants were presented with the first word and a 30-second-long presentation of the stationary starting frame, allowing participants to plan their subsequent animation of that word. Finally, individuals were given 45 seconds to animate the given word. This process was repeated for the total of five words (mocking, seducing, surprising, following, fighting) and on each trial participants were given the option to discard and repeat their animations if they were unhappy with the result. Only the final animations were analyzed.

Stimulus Selection

Our procedure resulted in a total of 255 animations (51 for each word), recorded at a frame rate of 133 frames / second. Animations were then visually inspected for sufficient length and movement coverage of more than two quadrants of the screen. 53 animations failed these quality control checks. The final stimulus set comprised 202 animations (42 mocking, 38 seducing, 36 surprising, 44 following, 42 fighting).

Ratings Collection

Participants

Thirty-seven healthy volunteers (31 females, M [SD] age = 21.30 [2.68] years, range = 18-32 years) were recruited from the University of Birmingham Research Participation Scheme and gave written informed consent to participate in this study. Post-hoc power calculations based on an online application by Judd et al.[34] (https://jakewestfall.shinyapps.io/two_factor_power/) confirmed that this study had 91.2 % power to find an effect of size Cohen's d (d) = 0.4 for the main hypothesis (1). An a priori power analysis of the complete study was not performed due to the lack of applications available to estimate effect sizes for the present analyses (a mixed effects model with more than one fixed effect). Participants received either course credit or money (£8 per hour) for their participation. None of the participants had previously taken part in stimulus development.

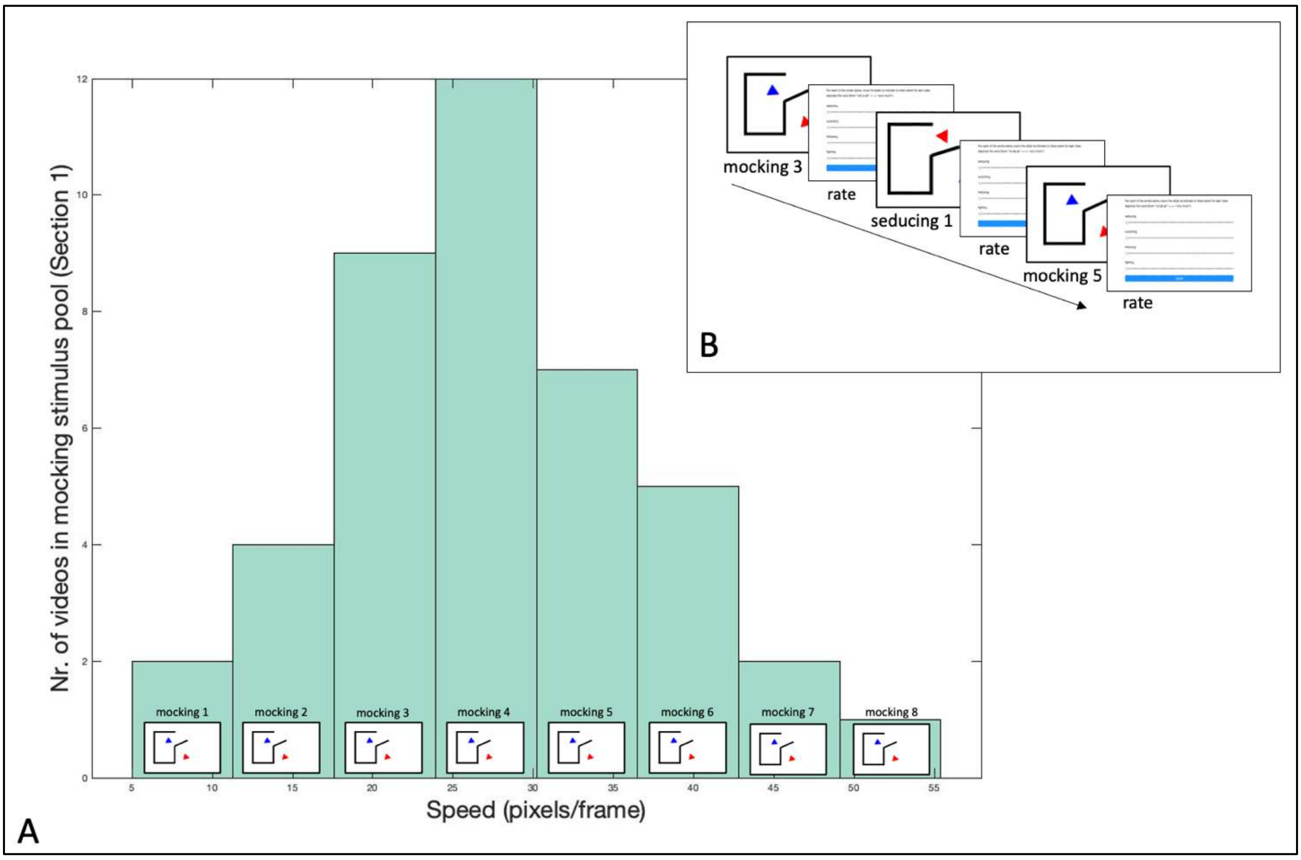
Task

The Ratings Collection phase comprised two tasks. First, all participants carried out a **production task**, where they created one 45-second-long animation for each of the five target words mocking, seducing, surprising, following and fighting, as described above. Following this, participants completed a **perception task**, where they viewed 40 animations from the full stimulus set and rated the extent to which the animations depicted each of the target words (mocking, seducing, surprising, following, fighting). Participants viewed eight exemplars of each of the five target words, presented in random order. The eight animations were selected from the stimulus pool (N = 202, see **Building the animotions database**) such that the mean speed of the triangles represented one of eight percentiles of the speed frequency distribution for a word (see Figure 5). Thus, for each word, each participant viewed a selection of animations such that they were exposed to the full range of kinematic variation in the population used to create the stimulus pool.

Finally, after watching each animation, participants were asked to rate on a visual analogue scale ranging from one to ten the extent to which they perceived the video to display the target word (e.g., mocking) and each of the four non-target words (e.g., seducing, surprising, following and fighting). The whole process of creating five and viewing and rating 40 45- second animations lasted between 40 and 50 minutes. Task order was fixed (production

Figure 5

Example of stimulus selection method.



Note. A) Example of the stimulus selection method for the word mocking. The selection method was the same for all five word categories. From each of eight percentile bins of the speed frequency distribution for a word category, one animation was selected at random and replayed to the participant. B) Schematic depiction of 3 successive trials in the perception task: Each animation was followed by a separate screen with five visual analogue sliding scales (one for each of the five word categories), ranging from 1 to 10.

task then perception task) to allow participants' animations to be unaffected by the animations they would see in the perception task. Due to the upper limit on the WACOM monitor refresh rate, videos were created with a 133 Hz sampling rate and displayed at 60Hz.

Procedure

Individuals sat in front of the WACOM Cintiq 22 HD touch screen, tilted at 30 degrees, and first completed a practice phase in which they familiarized themselves with moving the triangles around the screen. They were then instructed that they would first create an animation for each of the five words themselves (instructions were the same as in **Building the animotions database**; see Supplementary Materials) and subsequently would view and rate animations which had been created by other people. Participants then completed the production and perception tasks as described above.

Data Analysis and Processing

All data was processed in MATLAB R2020a[35] and analyzed in R[36]. Code required to reproduce data analysis and figures for this study will be freely available under (https://osf.io/pqn4u/).

Accuracy Ratings

Accuracy for each trial was calculated by subtracting the mean rating for all non-target words from the rating for the target word. Thus, a positive score indicates that the target word was rated higher than all non-target words, with higher accuracy scores reflecting better discrimination between target and non-target words. See Appendix 1 for further analysis of accuracy scores.

Spatial and Kinematic Predictors

All variables were calculated from positional data derived from the center points of the blue and red triangles. All steps of data processing mentioned below were performed on both the animations created by participants (= production data) and the animations from the full stimulus set used as perception task stimuli (= perception data).

Stimulus Kinematics

Instantaneous speed, acceleration magnitude and jerk magnitude were obtained by taking the first-, second- and third order non-null derivatives of the raw positional data, respectively (see [1], [2] and [3], where x and y represent x- and y positions of red and blue triangles in the cartesian coordinate system, v, a, and j denote instantaneous velocity, acceleration and jerk, respectively, and t denotes time).

$$\vec{v} = \sqrt{(x_{t-1} - x_t)^2 + (y_{t-1} - y_t)^2} \qquad [1]$$

$$\vec{a} = \frac{d\vec{v}}{d\vec{t}} \qquad [2]$$

$$\vec{j} = \frac{d\vec{a}}{d\vec{t}} \qquad [3]$$

As the ‘diff’ function in MATLAB amplifies the signal noise, which compounds for higher derivatives, we employed a smooth differential filter to undertake each step of differentiation (filter adopted from Huh & Sejnowski, 2015). The Euclidean norms of the x and y vectors of velocity, acceleration and jerk were then calculated to give speed, acceleration magnitude and jerk magnitude. That is, speed is calculated as the distance in pixels moved from one frame to the next. Acceleration magnitude comprises the change in speed from one frame to the next, and jerk magnitude comprises the change in acceleration. Mean speed, mean acceleration magnitude and mean jerk magnitude were then calculated by taking the mean across red and blue values, respectively. Lastly, kinematic values were converted from units of pixels/frame to mm/s.

Observer-Animator Kinematic Similarity

In order to measure the kinematic similarity between participants' and stimulus kinematics, absolute observer-animator jerk difference was calculated by first subtracting the mean jerk of each video a person rated from their own jerk values when animating the same word, and then taking the absolute magnitude of those values. Lower jerk difference values indicate *higher* observer-animator kinematic similarity.

Angular Frequency Spectral Density (AFSD)

For the purpose of quantifying animation trajectories, we adapted a method developed by Huh & Sejnowski (2015). Huh and Sejnowski have shown that the two-thirds power law varies according to shape trajectory, such that the gradient of the relationship between angular speed and curvature (in the Frenet-Serret frame[37,38]) is a function of the shape's angular frequency. Angular frequency here is defined as the number of curvature oscillations within one full tracing (360° or 2π radians) of a closed-form shape. We extended the method to derive the angular frequencies of arbitrary trajectories (i.e., not closed-form shapes) from the frequencies of speed oscillations within every 2π radians of a triangle's angular displacement in the Frenet-Serret frame.

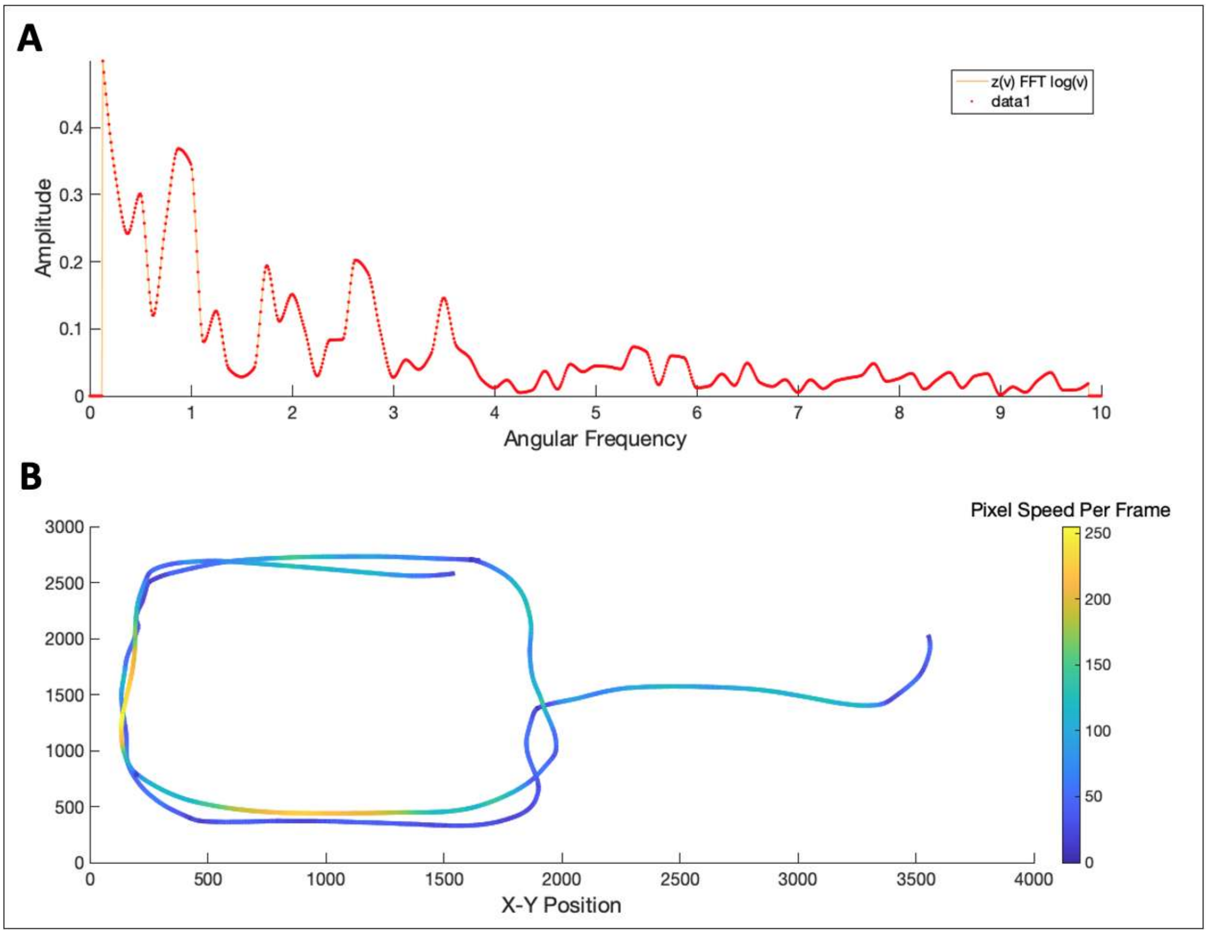
First, absolute instantaneous curvature k was calculated (angular speed divided by speed). This enables the calculation of Frenet-Serret speed v. Periodicity in v, in every 2π radians, allows the determination of angular frequencies present in the triangles' movement. Asymmetrical FFT was employed on log v, which returned the amplitude spectral density of all angular frequencies present for each triangle in each animation. Angular Frequency values returned by the FFT were then interpolated to obtain uniformly sampled values at 1001 points. Because the FFT assumes an infinite signal, when addressing a finite sample such as the log angular speed here, the first and last values of each sample must be continuous to avoid artefacts in the FFT results. We addressed this and any general drift in the signal (e.g., from participants generally slowing their movements due to fatigue) by removing a second order polynominal trend. The area under the amplitude spectral density curve was normalized to allow like to like comparison between differing lengths of red and blue triangle movement within and across participants. Across red and blue triangles' trajectories a weighted mean was then taken by multiplying each AFSD value with a factor reflecting the proportion of curved movement available for a triangle before averaging. See Figure 6 for an example of an amplitude spectrum and the related trajectory path.

Further Spatial Variables

A variety of other variables were created to further quantify spatial aspects potentially affecting inferences from the animations. First, simultaneous movement was calculated as the proportion of all frames where both red and blue triangles' speed was greater than zero (as seen in [4]), reflecting simultaneous movement of both triangles in a video. Furthermore, relative distance - the average distance between red and blue triangles - was quantified by taking the mean of the square root of the absolute distances between the triangles' x and y coordinates, respectively (see [5]). Finally, mean rotation reflects the average rotation of blue and red triangles around their own axis, measured in angle degrees and weighted by proportion of movement present across all frames for each color ([6]).

Figure 6

Example of trajectory shape and related angular frequency spectrum



Note. (A) Example of angular frequency spectrum for following animation. (B) Related trajectory (of one of two triangles). Trajectory colors indicate speed (pixel/frame).

$$\frac{\sum(speed_{red} \& \, speed_{blue} > 0.01)}{\sum all\ frames} \tag{4}$$

$$\underline{x}\left(\sqrt{\left(abs(x_{red} - x_{blue})\right)^2 + \left(abs(y_{red} - y_{blue})\right)^2}\right) \tag{5}$$

$$\frac{\left(\underline{x}\left(abs(r_{blue\ t-1} - r_{blue\ t})\right)\right) + \left(\underline{x}\left(abs(r_{red\ t-1} - r_{red\ t})\right)\right)}{2} \tag{6}$$

Statistical analysis

Data Analysis Overview

This study investigates the role of a large number of different predictor variables in explaining accuracy in the animations task. For two of these variables we present specific hypotheses (jerk, jerk difference); in addition, we wanted to investigate the role of a larger set of variables on an exploratory basis. For this reason, analyses were conducted in two stages: First, in a confirmatory stage, the roles of jerk and jerk difference were examined using Bayesian mixed models. Second, in an exploratory stage, a random forest model was performed, investigating the relative contribution of all predictor variables together.

Data Cleaning and Transformations

For all analyses, variables that were not normally distributed were either log- or square-root transformed to approximate normal distribution. Any outliers, as defined by values exceeding three scaled absolute deviations from the median, were replaced with the respective lower and upper threshold values. Finally, all predictor variables were z-scored.

Confirmatory analysis

A Bayesian linear mixed effects model was fitted in R using the *brms* package[39] to evaluate the relative contribution of jerk and jerk difference to accuracy as a function of word category, as well as their three-way interaction. A maximal[20] random effects structure was defined, allowing the intercept, the predictors of interest and their interactions to vary by participants (subject ID) and items (animation ID). Jerk and jerk difference were entered as covariates and word category was entered as dummy coded factor. We used brms default priors for the intercept and the standard deviation of the likelihood function as well as weakly informative priors (following a normal distribution centered at 0 and SD = 10) for all regression coefficients. Each model was run for four sampling chains with 5000 iterations each (including 1000 warmup iterations). There was no indication of convergence issues for any of the models (all Rhat values = 1.00, no divergent transitions).

Exploratory analysis I

Bootstrapped F-tests were performed to test for differences, between the five target words, in the presence of angular frequencies at each of the 1001 points on the amplitude spectrum. Bootstrapping amplitude spectrum values 1000 times revealed nine significant clusters, defined as clusters of difference that occurred in less than 5% of comparisons with resampled distributions (see Fig. 3A). The maxima and minima of each significant cluster were then used as bin edges for calculating the amplitude spectral density as the area under the curve within those nine bins, for both red and blue triangles' trajectories in each animation (cluster bin edges: 0.21 – 1.49, 1.61 – 2.39, 2.64 – 2.87, 3.04 – 3.40, 3.91 – 4.27, 4.79-5.19, 6.19-6.68, 7.6-7.93, 8.75-10). Finally, the weighted mean (weighted by amount of curved movement present in a triangle's full trajectory) was taken across red and blue triangles' spectral density values to form a single value of mean AFSD for each of nine bins for each animation. The final spectral density values are reflective of the relative proportion of curved movement available in a video in each of the nine areas of interest. Thus, a video that had high spectral density in bin 3 would be dominated by shapes with angular frequencies between 2.64 and 2.87. That is, relative to all other animations, the triangles in this video would be predominately moving with a speed and acceleration profile that lies between that of elliptical- and triangle trajectories.

Exploratory analysis II

Relative variable importance of 16 variables in predicting accuracy was assessed using random forest models[23] and the feature selection wrapper algorithm *Boruta*[24]. Random forests are ensembles of decision trees, where each tree is grown from a pre-specified subset of bootstrapped samples and where, for each tree, only a randomly selected subset of variables are considered as splitting variables. Boruta makes use of the *ranger* package[40] to train a random forest regression model on all variables as well as their permuted copies - so called "shadow features". First, *normalized permutation importance* (scaled by standard error, see[23]) of all features is assessed. Permutation importance of a given variable is the reduction in prediction accuracy (mean decrease in accuracy, MDA) of the model when this variable is randomly permuted. A variable is then classed as important when the Z-score of their importance measure is significantly higher than the highest importance Z-score achieved by a shadow feature. Overall performance of the model was evaluated by fitting a random forest with the ranger package with 500 trees and 10 random variables per tree.

Due to the known correlational structure within the data and the present lack of random forest models which can account for random effects, this analysis was performed items-based. For this purpose, for every variable, values corresponding to the same item were averaged across subjects, resulting in a total of 202 data points. Note that, due to the reliance on between-subject variance of variables relating to own-stimulus kinematic difference, these variables were excluded from this analysis.

References

1 Heider, F. & Simmel, M. An experimental study of apparent behavior. *The American Journal of Psychology* **57**, 243-259, doi:10.2307/1416950 (1944).

2 Abell, F., Happé, F. & Frith, U. Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development* **15**, 1-16, doi:https://doi.org/10.1016/S0885-2014(00)00014-9 (2000).

3 Klin, A. Attributing Social Meaning to Ambiguous Visual Stimuli in Higher-functioning Autism and Asperger Syndrome: The Social Attribution Task. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* **41**, 831-846, doi:10.1111/1469-7610.00671 (2000).

4 Castelli, F., Frith, C., Happé, F. & Frith, U. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* **125**, 1839-1849, doi:10.1093/brain/awf189 (2002).

5 Bliksted, V., Ubukata, S. & Koelkebeck, K. Discriminating autism spectrum disorders from schizophrenia by investigation of mental state attribution on an on-line mentalizing task: A review and meta-analysis. *Schizophrenia Research* **171**, 16-26, doi:https://doi.org/10.1016/j.schres.2016.01.037 (2016).

6 Newbury-Helps, J. *Are difficulties in mentalizing associated with severity of Antisocial Personality Disorder?* Ph.D. thesis, University College London, (2011).

7 Eddy, C. M. & Rickards, H. E. Interaction without intent: the shape of the social world in Huntington's disease. *Social Cognitive and Affective Neuroscience* **10**, 1228-1235, doi:10.1093/scan/nsv012 (2015).

8 Eddy, C. M. & Cavanna, A. E. Triangles, tricks and tics: Hyper-mentalizing in response to animated shapes in Tourette syndrome. *Cortex* **71**, 68-75, doi:10.1016/j.cortex.2015.06.003 (2015).

9 Edey, R. *et al.* Interaction takes two: Typical adults exhibit mind-blindness towards those with autism spectrum disorder. *Journal of Abnormal Psychology* **125**, 879-885, doi:10.1037/abn0000199 (2016).

10 White, S. J., Coniston, D., Rogers, R. & Frith, U. Developing the Frith-Happé animations: A quick and objective test of Theory of Mind for adults with autism. *Autism Research* **4**, 149-154, doi:https://doi.org/10.1002/aur.174 (2011).

11 Kenny, L. *et al.* Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism* **20**, 442-462, doi:10.1177/1362361315588200 (2016).

12 De Marco, D., Scalona, E., Bazzini, M. C., Avanzini, P. & Fabbri-Destro, M. Observer-Agent Kinematic Similarity Facilitates Action Intention Decoding. *Scientific Reports* **10**, 2605-2605, doi:10.1038/s41598-020-59176-z (2020).

13 Cook, R., Bird, G., Catmur, C., Press, C. & Heyes, C. Mirror neurons: from origin to function. *Behav Brain Sci* **37**, 177-192, doi:10.1017/s0140525x13000903 (2014).

14 Friston, K., Mattout, J. & Kilner, J. Action understanding and active inference. *Biological Cybernetics* **104**, 137-160, doi:10.1007/s00422-011-0424-z (2011).

15 Kilner, J., Hamilton, A. F. d. C. & Blakemore, S.-J. Interference effect of observed human movement on action is due to velocity profile of biological motion. *Social Neuroscience* **2**, 158-166, doi:10.1080/17470910701428190 (2007).

16 Rizzolatti, G. & Craighero, L. The mirror-neuron system. *Annu Rev Neurosci* **27**, 169-192, doi:10.1146/annurev.neuro.27.070203.144230 (2004).

17 Blythe, P. W., Todd, P. M. & Miller, G. F. *Evolution and cognition.* 257-285 (Oxford University Press, 1999).

18 Roux, P., Passerieux, C. & Ramus, F. Kinematics matters: A new eye-tracking investigation of animated triangles. *Quarterly Journal of Experimental Psychology* **66**, 229-244, doi:10.1080/17470218.2012.704052 (2013).

19 Huh, D. & Sejnowski, T. J. Spectrum of power laws for curved hand movements. *Proceedings of the National Academy of Sciences* **112**, E3950-E3958, doi:10.1073/pnas.1510208112 (2015).

20 Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* **68**, 10.1016/j.jml.2012.1011.1001-1010.1016/j.jml.2012.1011.1001, doi:10.1016/j.jml.2012.11.001 (2013).

21 Nalborczyk, L., Bürkner, P.-C. & Williams, D. R. Pragmatism should Not be a Substitute for Statistical Literacy, a Commentary on Albers, Kiers, and Van Ravenzwaaij (2018). *Collabra: Psychology* **5**, doi:10.1525/collabra.197 (2019).

22 Franke, M. & Roettger, T. *Bayesian regression modeling (for factorial designs): A tutorial* (2019).

23 Breiman, L. Random Forests. *Machine Learning* **45**, 5-32, doi:10.1023/A:1010933404324 (2001).

24 Kursa, M. B. & Rudnicki, W. R. Feature Selection with the Boruta Package. *2010* **36**, 13, doi:10.18637/jss.v036.i11 (2010).

25 Livingston, L. A., Shah, P. & Happé, F. *Further developing the Frith-Happé animations: A quicker, more objective, and web-based test of theory of mind for autistic and neurotypical adults* (2019).

26 Strobl, C., Malley, J. & Tutz, G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* **14**, 323-348, doi:10.1037/a0016973 (2009).

27 Van der Hallen, R., Manning, C., Evers, K. & Wagemans, J. Global Motion Perception in Autism Spectrum Disorder: A Meta-Analysis. *Journal of Autism and Developmental Disorders* **49**, 4901-4918, doi:10.1007/s10803-019-04194-8 (2019).

28 Anzulewicz, A., Sobota, K. & Delafield-Butt, J. T. Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism. *Scientific Reports* **6**, 31107, doi:10.1038/srep31107 (2016).

29 Cook, J. L., Blakemore, S.-J. & Press, C. Atypical basic movement kinematics in autism spectrum conditions. *Brain* **136**, 2816-2824, doi:10.1093/brain/awt208 (2013).

30 Johnson, B. P. *et al.* Understanding macrographia in children with autism spectrum disorders. *Research in Developmental Disabilities* **34**, 2917-2926, doi:https://doi.org/10.1016/j.ridd.2013.06.003 (2013).

31 Cook, J. From movement kinematics to social cognition: the case of autism. *Philos Trans R Soc Lond B Biol Sci* **371**, 20150372, doi:10.1098/rstb.2015.0372 (2016).

32 Gordon, A. S. & Roemmele, M. 49-60 (Springer International Publishing).

33 World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* **310**, 2191-2194, doi:10.1001/jama.2013.281053 (2013).

34 Judd, C. M., Westfall, J. & Kenny, D. A. Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology* **68**, 601-625, doi:10.1146/annurev-psych-122414-033702 (2017).

35 Matlab v. R2020a (Natick, Massachusetts, 2020).

36 R: A language and environment for statistical computing. v. 4.0.2 (Vienna, Austria, 2020).

37 Frenet, F. Sur les courbes à double courbure. *Journal de Mathématiques Pures et Appliquées*, 437-447 (1852).

38 Serret, J. A. Sur quelques formules relatives à la théorie des courbes à double courbure. *Journal de Mathématiques Pures et Appliquées* **16**, 193-207 (1851).
39 Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *2017* **80**, 28, doi:10.18637/jss.v080.i01 (2017).
40 Wright, M. N. & Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *2017* **77**, 17, doi:10.18637/jss.v077.i01 (2017).

Author contributions statement:

B.S., S.S. and J.C. conceived the experiments, B.S. and S.S. conducted the experiments, D.F. and D.H. contributed analysis tools. A.G. and J.v.d.B. contributed the code for the online task. B.S. analysed the results. B.S. and J.C. wrote the manuscript.

Additional Information:

The authors declare no competing interests.