



OPEN

Keeping in time with social and non-social stimuli: Synchronisation with auditory, visual, and audio-visual cues

Juliane J. Honisch^{1✉}, Prasannaheet Mane¹, Ofer Golan² & Bhisadev Chakrabarti^{1✉}

Everyday social interactions require us to closely monitor, predict, and synchronise our movements with those of an interacting partner. Experimental studies of social synchrony typically examine the social-cognitive outcomes associated with synchrony, such as affiliation. On the other hand, research on the sensorimotor aspects of synchronisation generally uses non-social stimuli (e.g. a moving dot). To date, the differences in sensorimotor aspects of synchronisation to social compared to non-social stimuli remain largely unknown. The present study aims to address this gap using a verbal response paradigm where participants were asked to synchronise a 'ba' response in time with social and non-social stimuli, which were presented auditorily, visually, or audio-visually combined. For social stimuli a video/audio recording of an actor performing the same verbal 'ba' response was presented, whereas for non-social stimuli a moving dot, an auditory metronome or both combined were presented. The impact of autistic traits on participants' synchronisation performance was examined using the Autism Spectrum Quotient (AQ). Our results revealed more accurate synchronisation for social compared to non-social stimuli, suggesting that greater familiarity with and motivation in attending to social stimuli may enhance our ability to better predict and synchronise with them. Individuals with fewer autistic traits demonstrated greater social learning, as indexed through an improvement in synchronisation performance to social vs non-social stimuli across the experiment.

Everyday social interactions such as talking with friends, playing games (i.e. pat a cake) or lifting a table together require us to continuously monitor and predict the behaviour of others, and adjust our movements accordingly to achieve smooth interactions. The ability to closely synchronise our movements with external stimuli has been proposed to have evolutionary functions for resource acquisition and reproduction¹ by facilitating cooperation² and social cohesion^{3,4}. One of the most common channels used for synchronous interactions is the verbal channel, as in simple parent-child games, singing, sloganeering or through the entrainment of speech rhythms during everyday conversations^{5,6}.

Previous research on sensorimotor synchronisation has largely used finger tapping paradigms to examine synchronous responses to non-social stimuli, such as flashing lights or auditory metronomes^{7,8}. However, most stimuli that we synchronise to are social in nature, e.g. the movements of another person, dance, speech or songs. Studies that have investigated synchronising to social stimuli have generally used a different class of paradigms with more ecological validity, e.g. synchronous arm curling⁹, walking together¹⁰, dance-like movements¹¹ or synchronous bouncing¹². The focus of these studies has been on social affiliative and cognitive outcomes rather than the accuracy of synchronous responses to social stimuli. Due to this historical diversity of experimental approaches in studying synchrony in social and non-social domains, the differences in synchronising to social vs non-social stimuli remain largely unknown. While social compared to non-social stimuli have been shown to be processed differently in multiple studies¹³⁻¹⁷, none of these studies have used synchrony based measures to examine how we physically interact with such stimuli. Most studies in this area use paradigms where observers passively observe social and non-social pictures/videos presented on a screen^{14,16-19}. In contrast, interpersonal synchrony involves an active involvement with the environment, and is more typical of real-world social situations where participants are not merely passive observers.

¹Centre for Autism, School of Psychology and Clinical Language Sciences, University of Reading, Reading RG6 6AL, UK. ²Department of Psychology, Bar-Ilan University, Ramat Gan, Israel. ✉email: J.J.Honisch@reading.ac.uk; b.chakrabarti@reading.ac.uk

Fixed effect omnibus tests				
	F	Num df	Den df	p
Condition	16.697	2	2590.7	<.001
Stimulus type	797.117	1	2604.7	<.001
AQ	0.166	1	40.1	0.686
Gender	0.307	1	39.9	0.583
Condition * Stimulus Type	44.432	2	2592.6	<.001

Table 1. Parameter estimates for the fixed effect omnibus tests.

The real world rarely offers stimuli in a single sensory modality. Accordingly, it is vital to consider cues in single as well as multiple sensory modalities within each class of stimuli (social and non-social). Previous reports using non-social stimuli have shown closer synchronisation performance in terms of mean asynchrony when synchronising to an auditory compared to a visual metronome^{20–22}. This finding is not unexpected, as audition has been found to be superior in temporal processing, whereas the visual modality is superior in spatial processing²⁰. However, visual dynamic stimuli compared to discrete visual stimuli have been found to improve synchronisation performance²³. Specifically, in bimodal (audio-visual) stimuli, individuals have been shown to optimally integrate information from both modalities²⁴. Elliott, Wing and Welchman²⁵ showed that synchronous finger tapping with bimodal discrete stimuli was more accurate compared with unimodal stimuli. However, little is known about the impact of the number of modalities on synchronisation to social stimuli.

Engaging in synchronous interactions facilitates social bonds, by increasing liking³, trust²⁶, prosocial behaviour^{27,28} and reducing outgroup bias²⁹. Individuals diagnosed with autism spectrum disorders (ASD) typically face challenges in social communication and sensory processing³⁰, and often experience difficulties in integrating multisensory information³¹. In lab-based studies, autistic individuals show a reduced preference for social compared to non-social stimuli^{32–34}. ASD has also been associated with atypical spontaneous facial mimicry and spontaneous motor synchronisation with another person^{35–38}. In comparison with typically developing children and adults, individuals with ASD have been found to produce weaker and more variable synchronisation behaviours to social stimuli^{39,40}.

The present study aims to systematically examine the differences in synchronisation behaviour to social compared to non-social stimuli. Here, we developed a novel verbal response paradigm in which participants synchronised their verbal response in time with social or non-social stimuli (presented unimodally, audio or visual) or bimodally (audio-visual). We predict that individuals will perform differently in how well they synchronise with social compared to non-social stimuli. One possibility is that individuals will be worse in synchronising to social stimuli due to their greater complexity (i.e., a face is significantly more visually complex compared to a dot). Another possibility is for individuals to show better performance in synchronising with social stimuli, due to the greater reward value typically attributed to social stimuli^{16,19}. The second possibility is supported by empirical and theoretical accounts suggesting greater reward response associated with motor alignment^{41–43}. For both social and non-social stimuli, synchronisation with audio-visual combined conditions is expected to result in lower asynchrony compared to the audio and visual only conditions²⁵. Further, in line with the literature on auditory superiority when synchronising with a metronome, we predict closer synchronisation with auditory compared with visual stimuli^{20–22}.

In this study, autistic traits were measured using the Autism Spectrum Quotient⁴⁴ (AQ). Autistic traits are distributed continuously throughout the population with similar aetiology at both ends, allowing us to examine the impact of autism-related variation at a population level⁴⁵. We predict that autistic traits will be negatively associated with synchronisation performances across tasks, in line with earlier findings of weaker and more variable synchronisation performances in individuals with ASD^{38–40}.

Results

Model fit statistics for the linear mixed model implemented were as follows:

$$\text{AIC} = -1726.84, \text{BIC} = -1602.964, \text{R-squared conditional} = 0.273$$

Results of the linear mixed model analysis are presented below in Tables 1 and 2.

The model revealed a main effect of Stimulus Type ($F(1,2604.7) = 797.12, p < 0.001$) with social stimuli associated with significantly lower mean absolute asynchrony (estimated marginal mean = 142 ms, 95%CI [129, 154]) compared to non-social stimuli (estimated marginal mean = 334 ms, 95%CI [322, 345]). A main effect of Condition was also observed ($F(2,2590.7) = 16.70, p < 0.001$) with visual stimuli (estimated marginal mean = 265 ms, 95%CI[251, 279]) associated with significantly larger absolute mean asynchrony compared with auditory (estimated marginal mean = 230 ms, 95%CI[216, 243]) and audio-visual combined stimuli (estimated marginal mean = 219 ms, 95%CI[216, 243], see Fig. 1). No significant difference between auditory and audio-visual combined conditions was observed. An interaction effect between condition and stimulus type was noted ($F(2,2592.6) = 44.43, p < 0.001$). All pairwise post-hoc comparisons show that social stimuli were associated with better synchronization performance compared to non-social stimuli, irrespective of sensory modality. For non-social stimuli, participants synchronized better with visual than with auditory conditions [$t(2600) = 3.160, p = 0.023$]. For social stimuli, synchronisation performance was better for auditory and audio-visual conditions in

Fixed effects parameter estimates								
Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	0.34766	0.0180	0.31248	0.38285	55.0	19.368	<.001
Condition1	C-A	-0.01649	0.0111	-0.03831	0.00533	2595.1	-1.481	0.139
Condition2	V-A	-0.03573	0.0113	-0.05788	-0.01357	2600.9	-3.161	0.002
Stimulus type 1	Social-Nonsocial	-0.24283	0.0117	-0.26571	-0.21995	2592.0	-20.804	<.001
AQ	AQ	3.96e-4	9.74e-4	-0.00151	0.00231	40.1	0.407	0.686
Gender 1	Male-Female	-0.00565	0.0102	-0.02564	0.01433	39.9	-0.554	0.583
Condition 1 * Stimulus type 1	C-A * Social-Nonsocial	0.01112	0.0164	-0.02111	0.04336	2587.5	0.677	0.499
Condition 2 * Stimulus type 1	V-A* Social-Nonsocial	0.14213	0.0167	0.10937	0.17488	2594.0	8.505	<.001

Table 2. Parameter estimates for the fixed effects of the model. Conditions are denoted as follows (A: Auditory, C: Combined, V: Visual). Stimulus Types are denoted as Social or Nonsocial.

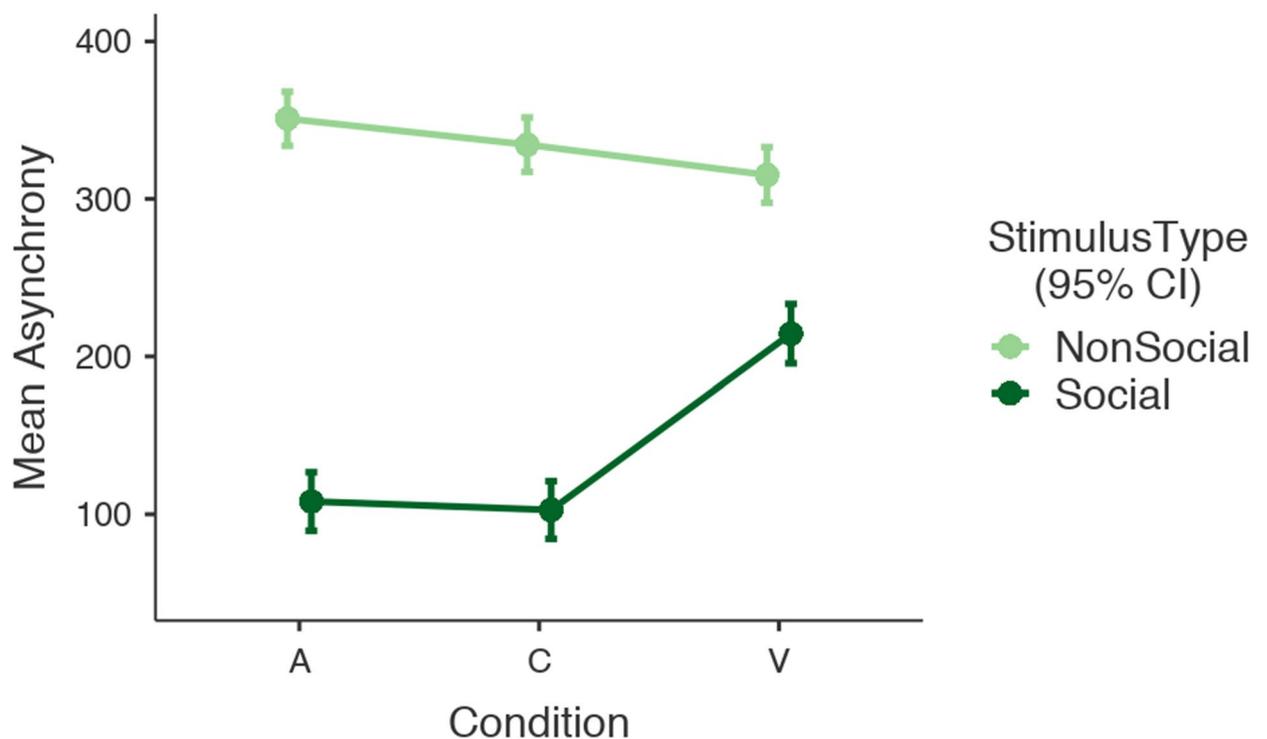


Figure 1. This figure shows the effect of Stimulus Type (Social, Non-social) and Condition (A: Auditory, C: Combined, V: Visual) on mean asynchrony in seconds. The error bars signify a 95% confidence interval.

comparison to visual conditions [$t_{\text{visual-audiovisual}}(2584) = 9.156, p < 0.001$; $t_{\text{auditory-visual}}(2584) = -8.644, p < 0.001$]. A full list of post-hoc comparisons is provided in the Supplementary Materials.

No effects of AQ ($F(1,40.1) = 0.166, p = 0.686$) or gender ($F(1,39.9) = 0.307, p = 0.583$) were noted.

It is possible to measure synchrony using either the visual or auditory onsets in response to the combined audio-visual stimulus. Synchronisation performance to social combined (audio-visual) stimuli was further examined to test if it differed as a function of using auditory vs visual onsets to define the target cue. This analysis was restricted to social stimuli alone, since auditory and visual onsets for non-social audio-visual stimuli were programmed to be identical. The results revealed a significant difference in mean asynchrony as a function of onset type [$t(2282) = -10.94, p_{\text{Bonferroni}} < 0.001$], with higher mean asynchrony associated with the visual onset (estimated marginal mean = 201 ms, 95% CI [264.1,320]) than with the auditory onset (estimated marginal mean = 142 ms, 95% CI [75.4,130]).

An exploratory analysis to measure learning/practice effects through the experiment revealed a main effect of trial [$F(1,2586.2) = 9.942, p = 0.002$], a significant two-way interaction of trial and stimulus type [$F(1,2586.2) = 7.341, p = 0.007$] as well as a significant three-way interaction between stimulus type, trial, and AQ [$F(2,2588.3) = 3.154, p = 0.043$] (see Figs. 2 and 3).

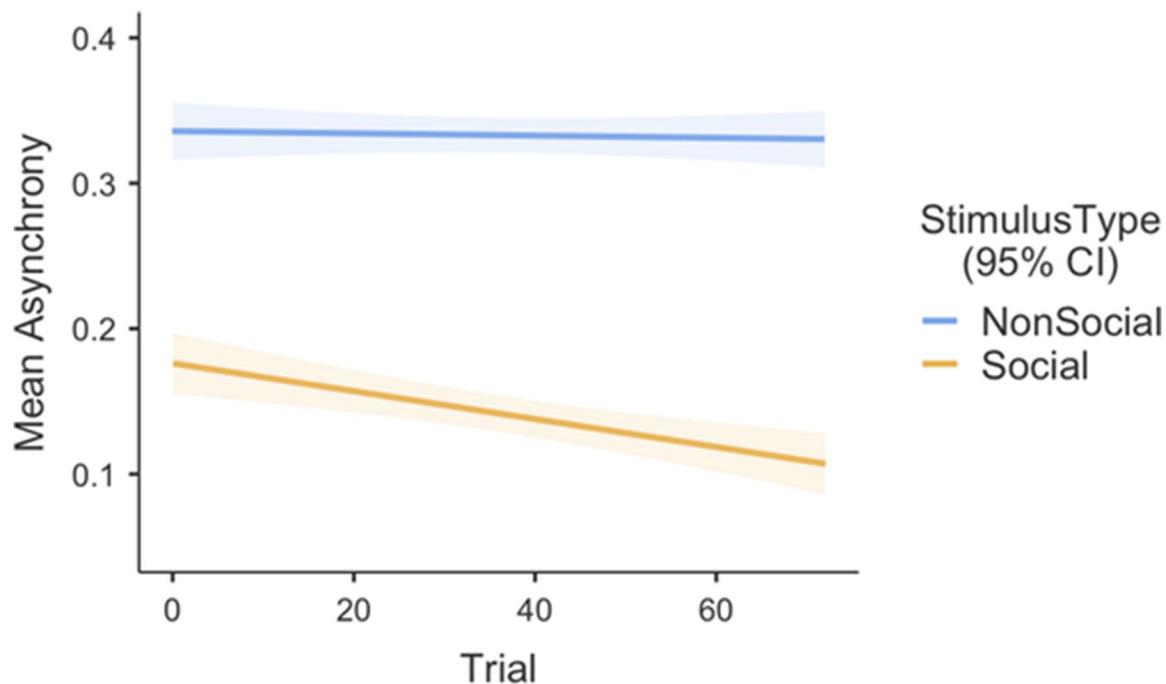


Figure 2. This figure shows the two-way interaction for the learning effect, indicating a decrease in mean absolute asynchrony for social but not for non-social conditions. Stimulus Type (Social, Non-social) on mean asynchrony in seconds. The error bars signify a 95% confidence interval.

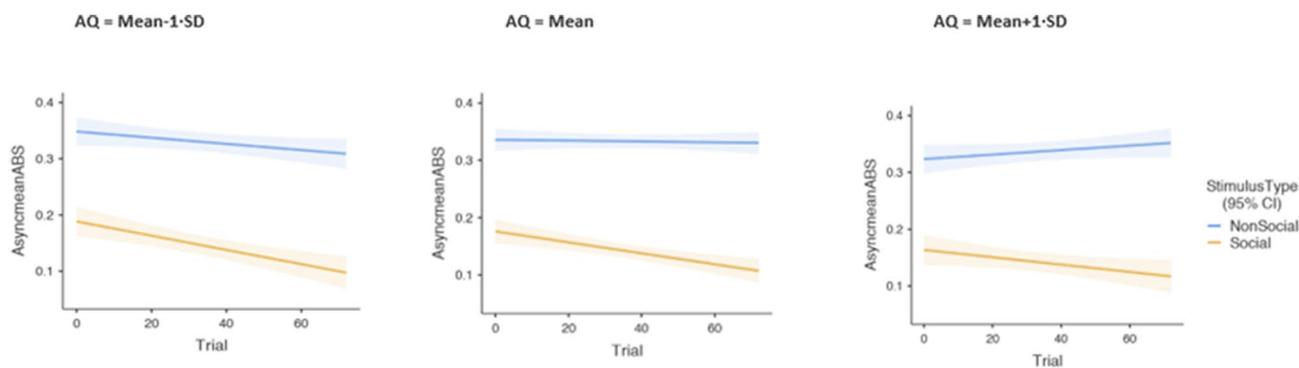


Figure 3. This figure shows the three-way interaction for the learning effect and AQ, indicating a decrease in mean absolute asynchrony across trials for social compared with non-social stimuli. The *left panel figure* illustrates that participants who scored 1 SD below the sample mean showed a steep decrease in mean absolute asynchrony across trials for social compared with non-social stimuli. In contrast, participants who scored 1 SD above the mean AQ scores (*right panel figure*) showed a less steep decrease in mean absolute asynchrony in response to social vs non-social stimuli across trials.

Figure 2 illustrates the data from the two-way interaction by showing a monotonically decreasing trend of mean absolute asynchrony for social trials, suggesting the existence of a practice/learning effect. This trend is not seen in response to the non-social trials. Figure 3 demonstrates the three-way interaction between stimulus type, trial, and AQ.

Discussion

The present study systematically tested the differences in synchronising with social and non-social stimuli, when presented in single or multiple sensory modalities. Mean asynchrony for trials in response to non-social stimuli was found to be greater than in response to social stimuli. Additionally, mean absolute asynchrony for visual stimuli were found to be higher than for auditory and audio-visual (combined) stimuli. No effect of autistic traits or gender were noted on mean absolute asynchrony.

Participants were significantly better in synchronising with a social compared to non-social stimuli, across all three conditions (auditory/visual/audio-visual), despite the greater variability in the target cue timings for social stimuli. This finding could potentially be an outcome of increased attention or reward value ascribed to social over non-social stimuli^{17,19,34,46}. Infants less than four weeks old have been found to show preferential

attention to social stimuli such as human sounds, movements and facial features^{47,48}. Increased reward value of target stimuli can influence both the starting point as well as the rate of sensory data accumulation^{49,50}. Greater relative reward value for social stimuli can therefore potentially explain closer synchronous performances for such stimuli as observed in our current study. Another plausible explanation for this difference is the use of verbal responses in the present study. Verbal responses match the nature of the social but not the non-social stimuli in our paradigm. The compatibility between the action performed in the social stimuli and the motor response performed by participants may have improved motor simulation by activating the same neural systems involved in perceiving and producing the same motor response⁵¹. Consequently, participants' predictive model of the stimuli, along with participants' motor planning and execution could have been enhanced, resulting in closer synchronous performances⁵². Further, given the social nature of human actions, whether it is making a *ba* sound or a finger tap, it is not entirely possible to decouple response compatibility from the sociality of stimuli. Another possible contributing factor for the improved synchronisation observed for social stimuli could be the lower pitch of the male and female actor in comparison with the higher pitch of the metronome. Synchronisation with auditory metronomes can be influenced by the pitch of the sounds, with better synchronisation associated with lower pitch sounds^{53,54}. However, no systematic analysis has been conducted to estimate the magnitude of this effect on a wide range of pitches that are comparable to those used in this study, and whether this finding can be generalised to human vocal sounds.

Our exploratory analysis to check for practice/learning effects revealed a significant difference between the social and non-social conditions. Across the whole sample, individuals tended to perform better (i.e., with less absolute mean asynchrony) as the experiment progressed. This learning effect was greater for social compared to non-social stimuli, which could reflect greater attention being drawn to the dynamic social stimulus (a face saying 'ba ba') in comparison to the bouncing dot. This social learning effect was greater in individuals with lower autistic traits, than those with higher autistic traits. This observation is consistent with a recent report showing reduced integration of social information in a learning task in individuals with high autistic traits⁵⁵.

The observation of differences in response to social and non-social stimuli raises an important question about the origin of these differences. The low-level properties of social and non-social stimuli used in this study are different (e.g., contrast, colour, nature of sound). It is possible that the distinction between 'social' and 'non-social' stimuli is a cumulative effect driven by a large number of such low-level properties. Perfect matching for all stimulus properties will render the two sets of stimuli identical to one another. It is worth noting that the category of 'social' stimuli represents a circumscribed set of low-level stimulus features (e.g., flesh tones for visual stimuli, sound within the vocal frequency range for auditory stimuli). Whether there is a 'social' advantage over and above all potential physical characteristics of the stimulus remains an open question. Although our paradigm has higher ecological validity by presenting a real recording of a human partner, examining synchronisation that replicates a real-life social interaction that involves mutual and reciprocal adaptation would be of interest for future research.

In view of the significant interaction between stimulus type and condition observed in the main analysis, we separately analysed synchronisation performance by stimulus type. For non-social stimuli, there was little difference between the three conditions (see Fig. 1), but participants synchronised better with visual, than with auditory stimuli. This finding is in contrast to a previous study that showed better synchronization of finger tap responses to an auditory compared to a visual stimulus⁵⁶. One potential explanation is that continuous visual stimuli provide a more salient temporal cue to allow for better temporal judgements than discrete auditory stimuli⁵⁷. The continuity of visual stimuli like the ones used in this study provide sufficient time to the participant to prepare a response to synchronize with the target cue^{58,59}. Synchronization performance with the combined audio-visual stimulus was not significantly different from that with either of the unimodal stimuli.

For social stimuli, an opposite pattern was observed—with participants synchronising significantly better with auditory than visual stimuli. This pattern of results is consistent with the observation that the auditory modality performs better than vision in tasks that involve temporal processing^{60,61}. One possibility is that this pattern of auditory dominance in synchronisation tasks is observed once the mismatch between the target stimulus and response modality is minimised. For finger tapping tasks such as the one by Hove and others⁶², the visual stimulus of a moving bar was similar to the response modality (a moving finger). Once this mismatch is minimised, the cognitive efforts could be directed entirely to the temporal aspects of the stimuli, which would result in an auditory dominance effect. In the current study, the mismatch between the target stimulus and the response modality is significantly lower for the social stimuli than the non-social stimuli, which leads to the expected pattern of auditory dominance for the social stimuli. For the main analysis, all target cues were identified auditorily, i.e. the time of peak of every 'ba' utterance was used to calculate the asynchrony. When visual onsets (first frame of mouth opening) were used to calculate the asynchrony, an identical pattern was observed for synchronisation with unimodal auditory and visual stimuli. Interestingly however, the synchronisation performance with the combined audio-visual stimulus differed significantly as a function of which onset type was used. When auditory onsets were used, synchronisation performance was similar to the unimodal auditory condition. However, when the visual onsets were used, performance was closer to, but significantly worse than the unimodal visual condition. This result suggests that participants tend to synchronise their verbal responses to the auditory rather than the visual cue in the audio-visual condition.

Contrary to our original hypothesis, no effect of autistic traits was noted in relation to synchronisation performance in response to either social or non-social stimuli per se. While this result is consistent with a recent report showing no autistic deficit in auditory-motor synchronisation using a finger tapping task⁶³, it is in contrast to another report using coherence as a measure of synchronization of bodily movements between a live experimenter and the participant⁴⁰. A direct comparison of the current results with these previous studies is not straightforward due to the different nature of stimuli and response modalities used. We note, however,

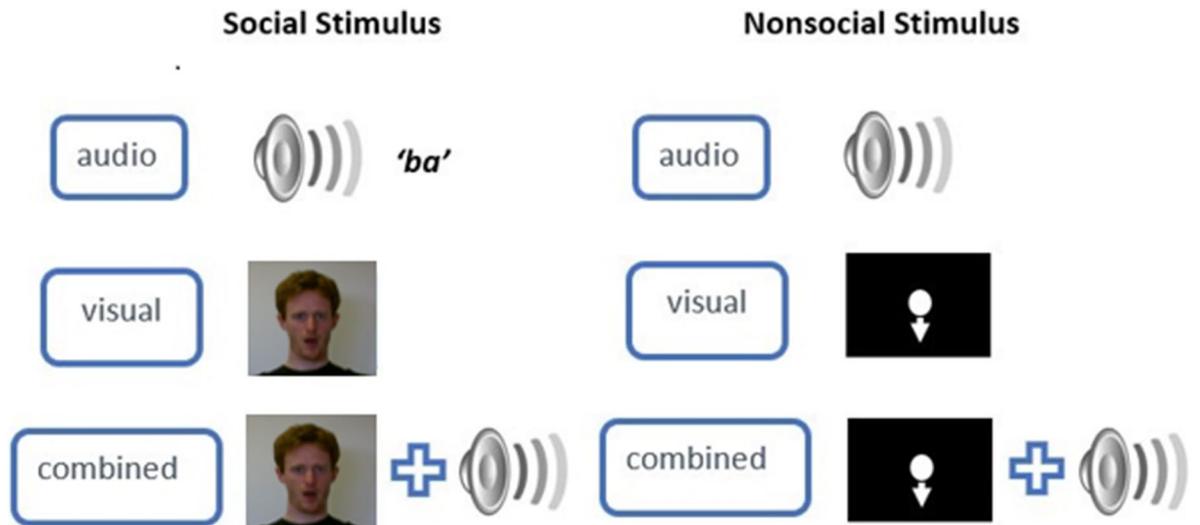


Figure 4. An illustration of the experimental design. Each stimulus type (social or non-social), contained audio only, visual only and audio-visual combined stimuli. For social stimuli, the audio condition consisted of a recording of a rhythmical verbal ‘ba’ response (either female or male), the visual condition consisted of a video recording of a female or male person performing the rhythmical ‘ba’ response without the presentation of the sound, and the combined condition consisted of a video recording of the female or male person performing the rhythmical ‘ba’ response with the sound. For non-social stimuli, either a rhythmical metronome beep (audio), a white vertically moving dot (visual) or both the metronome beep coinciding with the moving dot (combined) were presented.

that consistent with a previous report, a weaker social learning effect was observed in individuals with higher autistic traits⁵⁵.

In summary, our findings suggest that humans synchronise their responses more closely with social compared with non-social stimuli. This ‘social advantage’ is likely to be driven by the preferential attention and reward linked to perceiving and interacting with other humans. Potential future research could formally examine the dependence of these results on the response modality (verbal response/ finger tapping), as well as test the impact of attention given to social compared to non-social stimuli on synchronisation performance.

Materials and methods

Participants and design. Fifty-three psychology undergraduates (29 females, 24 males; $M_{\text{age}} = 21.01$ years, range = 18–31 years) took part in the study in exchange for course credit or for cash and were screened for photosensitive epilepsy. The study had a 2 (stimulus type: social, non-social) \times 3 (condition: audio, visual, audio-visual combined) within-subject design. Sample size was determined *a priori* using G*Power 3.1⁶⁴. The analysis was based on an effect size calculated from a previously published study demonstrating improved synchronous performance to auditory compared to visual cues²⁵ ($d = 1.3$). However, since our response modality as well as stimuli type differed from the study above, we chose a more conservative effect size ($d = 0.7$). The analysis suggested that the minimum acceptable total sample size needed to achieve a power of .80 was 41.

Ethics. The study was approved by the School of Psychology and Clinical Language Sciences Ethics Committee at the University of Reading. The experiment was performed in accordance with the relevant guidelines and regulations, and participants provided informed, written consent. Written informed consent was obtained to publish identifying images of the social stimulus, in an online open access publication.

Experiment setup and apparatus. Participants synchronised their verbal responses (a ‘ba’ sound) to either audio, visual or audio-visual combined target cues. A SparkFun sound detector and an acquisition hardware (National Instruments NI-USB 6343) were used to record the presence of participants’ verbal responses. One computer was used to control and present both the visual and auditory stimuli (target cues) on the screen.

Two types of target cues were presented, social and non-social (see Fig. 4). The social cues consisted of video recordings of both a male and a female actor performing a rhythmical ‘ba’ sound. The gender of the actor in the video recordings was matched with the gender of the participant, controlling for gender effects during synchronous activities⁵⁵.

The actor was presented from the collar bone upwards, with a neutral facial expression, a controlled background, and wearing fitted black clothes. For visual social cue conditions, the video recording was presented without sound. Here the opening of the actor’s mouth was the target cue. For audio social conditions, a blank black screen with a white fixation cross was presented with the audio recording of the actor’s rhythmical ‘ba’ sound. For the male recording the ‘ba’ sound was presented at an average of 110 Hz (minimum 108 Hz, maximum 112 Hz). Female ‘ba’ sounds were presented at an average of 211 Hz (minimum 207, maximum 215 Hz). In audio-visual social conditions, both the video and the corresponding audio recording were presented simultaneously.

Non-social cues consisted of inanimate stimuli. For visual non-social condition, a white dot (2.54 cm diameter) was presented on a black background. The refresh rate of the PC monitor was 60 Hz. The dot was moving vertically with a fixed amplitude of 20 cm, using a pre-generated sine wave function. The lowest point of the downwards motion was the target cue. In the auditory non-social conditions, the generated trajectory of the dot movement was used to estimate the corresponding non-social auditory cue. The lowest peak on the x-axis for each downwards oscillation was used to generate a series of rhythmical metronome tones with a tone duration of 50 ms and at 700 Hz. In audio-visual non-social conditions, both the dot motion with its corresponding audio metronome tones were presented simultaneously (see Supplementary Material for stimulus generation code). Although there could be small discrepancies under the 10-ms range between the programmed and presented audio/visual onsets for the non-social stimuli⁶⁶, we minimised the risk of such discrepancies by using a powerful graphics card (NVIDIA GTX 650, 4 GB), and a significantly higher screen refresh rate than the stimulus frequency. However, it should be noted that in the absence of an external photodiode to verify stimulus timings, it is not possible to quantify the magnitude of such discrepancies.

The rhythmical presentation of the target cues was varied across trials to minimise participants learning the tempo. Each trial for all conditions contained a tempo change to ensure that participants paid attention to the target cues. Six trials started with a fast, followed by a slow rhythm, and a further six trials followed the reverse order. The inter-target cue-intervals (ITI) for the fast tempo were on average 650 milliseconds ($\pm 5\%$), and 870 ms ($\pm 5\%$) for the slow tempo. The tempo change occurred randomly between the 5th or 7th ITI. Each condition contained 12 trials with each trial lasting 40 seconds, with an overall total of 72 trials (12 (trials) \times 6 (conditions: visual social, audio social, audio-visual social, visual non-social, audio non-social, audio-visual non-social)). The presentation of both the video stimuli and the generated dot motion was controlled by Psychophysics toolbox⁶⁷ in MATLAB (version 2014a; The Mathworks Inc., MA, USA).

The AQ was used to measure an individual's autistic traits⁴⁴. The AQ has 50 items measuring diverse dimensions of the autistic phenotype, such as, "I enjoy meeting new people". Participants rate their level of agreement with each statement on a 4-point Likert scale ranging from 'definitely agree' to 'definitely disagree'. Ratings are then collapsed to a yes/no scoring. Thus, the AQ scores range from 0 to 50, with autistic individuals typically scoring higher than neurotypicals. In the present study an online version of the questionnaire was administered.

Task and procedure. Participants were asked to attend two experimental sessions, each lasting around 50 minutes, with a minimum gap of one day from one another. Each session contained either the social or non-social stimuli. The order of these sessions was counterbalanced across participants. Participants completed the AQ online when signing up for this study. For the experimental sessions, participants arrived at the laboratory individually, were greeted by the experimenter, and were seated at a table facing a PC monitor. Participants learned that the goal of the experiment was to examine the effects of different types of stimuli on their ability to verbally synchronise with them. Once participants read the information, consent was provided. In the social-stimuli session, female participants were shown the video and audio recordings of the female actor, while male participants were shown the recordings of the male actor. For visual social conditions, participants were instructed to produce a 'ba' sound in synchrony with the mouth opening of the actor presented in the video. In auditory social conditions, participants were asked to synchronise their 'ba' response in time with the 'ba' sound of the actor. For audio-visual social conditions, participants were instructed to synchronise their responses to both the audio-visual cue ('ba' sound and mouth opening) of the actor presented in the video. In the non-social stimuli session, both male and female participants were either asked to produce a 'ba' sound in synchrony with computer-generated audio and visual stimuli. For non-social auditory conditions, participants synchronised their 'ba' response in time with an auditory metronome, whereas for non-social visual conditions participants synchronised their responses with a moving white dot at the lowest point on the vertical axis. Finally, for non-social audio-visual conditions, participants were instructed to synchronise their 'ba' responses with both the metronome and the moving dot simultaneously. The duration of each experimental session was 40 minutes.

Analysis. The synchrony analysis adopted an information-processing approach rather than a dynamical systems approach. The latter is more favoured by researchers who examine the continuous rhythmical movements. However, in the present study participants were instructed to synchronise their verbal response to match those of an external target, rather than simply to maintain a continuous rhythm. This information processing approach has been widely used by researchers who have examined synchronous performances between two or more individuals; for example, to analyse finger tapping⁶⁸, oscillatory arm movements⁶⁹, bouncing¹² and sound recording from a string quartet⁷⁰. For each trial, we recorded the sound onsets for all verbal responses performed by participants. We then used a custom-made peak detection algorithm in Matlab to extract the onset times for each verbal response that occurred after the tempo change. Response data before and at time of the period change were excluded from the data analysis to reduce additional variability introduced due to a different starting tempo and adjustments made to entrain to a new tempo. The alignment of target and response onsets was achieved in a way where the closest response onset to the target onset was used to estimate asynchrony. For the first response onset the target onset would always precede the response onset. Missing responses were interpolated adopting methods used in previous research⁷¹⁻⁷³. The following interpolation was conducted to account for missing responses; if a participant's response was two times as large as the target cue's tempo (inter-onset-interval, IOI), the participant's inter-response-interval (IRI) was split into two (divided by two). Similarly, if a participant's response was three times as large as the target cue's IOI, the participant's response was divided into three equal parts to account for the missing responses. Any responses larger than three times the relative target cue's IOI was discarded. Absolute asynchronies were calculated to indicate the magnitude of asynchrony, irrespective of a participant being ahead or behind the target stimulus^{74,75}.

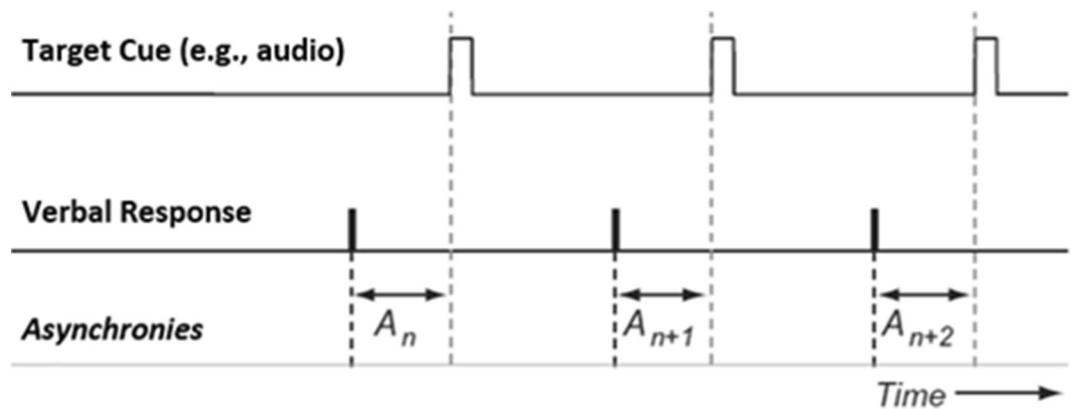


Figure 5. An illustration of the asynchrony calculation. Asynchrony is calculated as the difference between the event time of the respondent and the closest event time from the cue. Note, absolute asynchronies were used for the present analysis.

For non-social conditions, absolute asynchronies were calculated by calculating the difference between the target cue event time and the participant's corresponding response event time (see Fig. 5). Non-social target cue event times were taken from the stimulus file. For non-social audio-visual combined conditions, the target onsets for both the visual and auditory cue coincided in time. The target cue event times for visual-social conditions were estimated by two independent coders in Elan⁷⁶. The video recordings in the visual-social conditions were presented at a rate of 30 frames per second (30fps). Coders identified the first frame of mouth opening as the target event time. For audio-social conditions of the audio data from the videos was separated and saved as a wav file. The audio data was then smoothed using a bi-directional second order Butterworth low-pass filter⁷⁷. Maximum peaks were detected using an adaptive peak detector with a threshold of a valley preceding each maximum peak. Each audio target onset event was visually cross validated with a spectrogram of the raw signal⁷⁰. For social stimuli, it is possible to measure synchrony using either the visual or auditory onsets for the combined audio-visual conditions. We therefore examined both synchronisation performance with visual and auditory onsets for audio-visual combined conditions (see Results). However, synchronisation performance for audio-visual conditions has previously been reported to be comparable with that of the unimodal auditory conditions²⁰. Therefore the audio onsets, as extracted for the audio-social conditions, were used as the target cue for audio-visual condition for the primary analysis. To explore if the use of auditory vs visual onsets had a significant impact on asynchrony, we ran a separate analysis only on the social trials (model details in the following section).

Lastly, we examined the variability of the social stimulus onsets by computing the median standard deviation across all conditions (median 0.0347 s, minimum 0.0190 s, maximum 0.220 s).

Data reduction. Participants were excluded from the relevant analyses if one or more of these criteria were met: (a) being greater/less than 3 SD from the group mean ($N = 1$), and (b) missing data on a stimulus type for all conditions ($N = 8$). A linear mixed model analysis on the mean absolute asynchrony data was conducted from the remaining 42 participants. (see Table 1 for descriptives).

Statistical analyses. A linear mixed model implemented in jamovi v1.1.9⁷⁸ was defined to analyse the mean absolute asynchrony data, across all trials and after two response cycles following the tempo change in the stimulus. The model was as follows:

$$\text{Mean Absolute Asynchrony} \sim 1 + \text{Condition (Auditory/Visual/Audio - visual)} \\ + \text{Stimulus Type (Social/Non - social)} + \text{AQ} + \text{Gender} + \text{Condition} * \text{StimulusType} + (1|\text{Participant}).$$

Stimulus Type (social, non-social), Condition (audio, visual or combined), Gender, AQ scores were defined as fixed effects and participants were defined as random effects. Model fit was estimated using the Restricted Maximum Likelihood (REML) method.

To check if the synchronisation performance to social combined (audio-visual) stimuli differed as a function of using auditory vs visual onsets to define the target cue, a further analysis was conducted to compare the mean asynchrony estimated by the visual and the auditory onsets respectively. This analysis was done only for trials were social stimuli were presented, since auditory and visual onsets were identical for non-social audio-visual stimuli (as programmed). The following model was estimated:

$$\text{Mean Asynchrony} \sim 1 + \text{AQ} + \text{Condition (Auditory/Visual/Audiovisual)} \\ + \text{OnsetType (Auditory/Visual)} + \text{Condition} * \text{OnsetType} + (1|\text{Participant})$$

To explore the effect of learning/practice effects through the experiment, a separate model was tested, including Trial number as a predictor. The model is specified below:

$$\begin{aligned} \text{Mean Absolute Asynchrony} \sim & 1 + \text{Condition}(\text{Auditory/Visual/Audio} - \text{visual}) \\ & + \text{Stimulus Type} (\text{Social/Non} - \text{social}) + \text{AQ} + \text{Gender} + \text{Trial} + \text{Condition} * \text{StimulusType} \\ & + \text{StimulusType} * \text{Trial} + \text{StimulusType} * \text{Trial} * \text{AQ} + (1|\text{Participant}). \end{aligned}$$

Received: 20 May 2020; Accepted: 31 March 2021

Published online: 22 April 2021

References

1. Merker, B. Synchronous chorusing and the origins of music. *Music. Sci.* **3**, 59–73 (1999).
2. Wiltermuth, S. S. & Heath, C. Synchrony and cooperation. *Psychol. Sci.* **20**, 1–5 (2009).
3. Hove, M. J. & Risen, J. L. It's all in the timing: Interpersonal synchrony increases affiliation. *Soc. Cogn.* **27**, 949–960 (2009).
4. Miles, L. K., Nind, L. K. & Macrae, C. N. The rhythm of rapport: Interpersonal synchrony and social perception. *J. Exp. Soc. Psychol.* **45**, 585–589 (2009).
5. McGarva, A. R. & Warner, R. M. Attraction and social coordination: Mutual entrainment of vocal activity rhythms. *J. Psycholinguist. Res.* **32**, 335–354 (2003).
6. Wilson, M. & Wilson, T. P. An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* **12**, 957–968 (2005).
7. Repp, B. H. & Su, Y. H. Sensorimotor synchronization: A review of recent research (2006–2012). *Psychon. Bull. Rev.* **20**, 403–452 (2013).
8. Aschersleben, G. Temporal control of movements in sensorimotor synchronization. *Brain Cogn.* **48**, 66–79 (2002).
9. Miles, L. K., Nind, L. K., Henderson, Z. & Macrae, C. N. Moving memories: Behavioral synchrony and memory for self and others. *J. Exp. Soc. Psychol.* **46**, 457–460 (2010).
10. Wiltermuth, S. Synchrony and destructive obedience. *Soc. Infl.* **7**, 78–89 (2012).
11. Kirschner, S. & Tomasello, M. Joint drumming: Social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* **102**, 299–314 (2009).
12. Fraser, D.S., Honisch, J.J., E. M.T., Quinn, K.A. & Cacioppo, J. T. 6th Joint Action Meeting (International Conference). in *Contributions of fluency to the synchrony-liking relationship* 7 (2015).
13. Pfabigan, D. M., Gittenberger, M. & Lamm, C. Social dimension and complexity differentially influence brain responses during feedback processing. *Soc. Neurosci.* **14**, 26–40 (2019).
14. Falck-Ytter, T., Bakker, M. & Von Hofsten, C. Human infants orient to biological motion rather than audiovisual synchrony. *Neuropsychologia* **49**, 2131–2135 (2011).
15. Gray, K. L. H., Haffey, A., Mihaylova, H. L. & Chakrabarti, B. Lack of privileged access to awareness for rewarding social scenes in autism spectrum disorder. *J. Autism Dev. Disord.* **48**, 3311–3318 (2018).
16. Hedger, N., Haffey, A., McSorley, E. & Chakrabarti, B. Empathy modulates the temporal structure of social attention. *Proc. R. Soc. B Biol. Sci.* **285**, 20181716 (2018).
17. Pierce, K., Conant, D., Hazin, R., Stoner, R. & Desmond, J. Preference for geometric patterns early in life as a risk factor for autism. *Arch. Gen. Psychiatry* **68**, 101–109 (2011).
18. Elsabbagh, M. et al. The development of face orienting mechanisms in infants at-risk for autism. *Behavioural Brain Research* **251**, 147–154 (2013).
19. Chakrabarti, B., Haffey, A., Canzano, L., Taylor, C. P. & McSorley, E. Individual differences in responsivity to social rewards: Insights from two eye-tracking tasks. *PLoS ONE* **12**, e0185146 (2017).
20. Repp, B. H. & Penel, A. Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 1085–1099 (2002).
21. Repp, B. H. & Penel, A. Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychol. Res.* **68**, 252–270 (2004).
22. Kato, M. & Konishi, Y. Auditory dominance in the error correction process: A synchronized tapping study. *Brain Res.* **1084**, 115–122 (2006).
23. Hove, M. J., Fairhurst, M. T., Kotz, S. A. & Keller, P. E. Synchronizing with auditory and visual rhythms: An fMRI assessment of modality differences and modality appropriateness. *Neuroimage* **67**, 313–321 (2013).
24. Wing, A. M., Dumas, M. & Welchman, A. E. Combining multisensory temporal information for movement synchronisation. *Exp. Brain Res.* **200**, 277–282 (2010).
25. Elliott, M. T., Wing, A. M. & Welchman, A. E. Multisensory cues improve sensorimotor synchronisation. *Eur. J. Neurosci.* **31**, 1828–1835 (2010).
26. Launay, J., Dean, R. T. & Bailes, F. Synchronization Can Influence Trust Following Virtual Interaction. <https://doi.org/10.1027/1618-3169/a000173> (2013).
27. Valdesolo, P. & DeSteno, D. Synchrony and the social tuning of compassion. *Emotion* **11**, 262–266 (2011).
28. Cirelli, L. K., Einarson, K. M. & Trainor, L. J. Interpersonal synchrony increases prosocial behavior in infants. *Dev. Sci.* **17**, 1003–1011 (2014).
29. Inzlicht, M., Gutsell, J. N. & Legault, L. Mimicry reduces racial prejudice. *J. Exp. Soc. Psychol.* **48**, 361–365 (2012).
30. American Psychological Association (APA). <https://www.apa.org/>.
31. Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L. & Wallace, M. T. Altered auditory and multisensory temporal processing in autism spectrum disorders. *Front. Integr. Neurosci.* **4**, 129 (2011).
32. Chevallier, C., Kohls, G., Troiani, V., Brodtkin, E. S. & Schultz, R. T. The social motivation theory of autism. *Trends Cogn. Sci.* **16**, 231–239 (2012).
33. Chevallier, C. et al. Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Res.* **8**, 620–628 (2015).
34. Pierce, K. et al. Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. *Biol. Psychiatry* **79**, 657–666 (2016).
35. Hamilton, A.F.D. C. Emulation and mimicry for social interaction: A theoretical approach to imitation in autism. *Q. J. Exp. Psychol.* **61**, 101–115 (2008).
36. Beall, P. M., Moody, E. J., McIntosh, D. N., Hepburn, S. L. & Reed, C. L. Rapid facial reactions to emotional facial expressions in typically developing children and children with autism spectrum disorder. *J. Exp. Child Psychol.* **101**, 206–223 (2008).
37. McIntosh, D. N., Reichmann-Decker, A., Winkielman, P. & Wilbarger, J. L. When the social mirror breaks: Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Dev. Sci.* **9**, 295–302 (2006).

38. Georgescu, A. L. *et al.* Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: A motion energy study. *Mol. Autism* **11**, 11 (2020).
39. Fitzpatrick, P. *et al.* Impairments of social motor synchrony evident in autism spectrum disorder. *Front. Psychol.* **7**, 1323 (2016).
40. Fitzpatrick, P. *et al.* Social motor synchronization: insights for understanding social behavior in autism. *J. Autism Dev. Disord.* **47**, 2092–2107 (2017).
41. Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z. & Gvirts, H. Z. Herding brains: A core neural mechanism for social alignment. *Trends Cogn. Sci.* **23**, 174–186 (2019).
42. Hsu, C. T., Sims, T. & Chakrabarti, B. How mimicry influences the neural correlates of reward: An fMRI study. *Neuropsychologia* **116**, 61–67 (2018).
43. Hsu, C.-T., Neufeld, J. & Chakrabarti, B. Reduced reward-related neural response to mimicry in individuals with autism. *Eur. J. Neurosci.* **47**, 610–618 (2018).
44. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* **31**, 5–17 (2001).
45. Robinson, E. B. *et al.* Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch. Gen. Psychiatry* **68**, 1113–1121 (2011).
46. Libera, C. D. & Chelazzi, L. Learning to attend and to ignore is a matter of gains and losses. *Psychol. Sci.* **20**, 778–784 (2009).
47. Maurer, D. & Salapatek, P. Developmental changes in the scanning of faces by young infants. *Child Dev.* **47**, 523 (1976).
48. Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* **40**, 1–19 (1991).
49. Platt, M. L. & Glimcher, P. W. Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238 (1999).
50. Rorie, A. E., Gao, J., McClelland, J. L. & Newsome, W. T. Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the Macaque Monkey. *PLoS ONE* **5**, e9308 (2010).
51. Rizzolatti, G. & Craighero, L. The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192 (2004).
52. Knoblich, G. & Sebanz, N. Evolving intentions for social interaction: From entrainment to joint action. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 2021–2031 (2008).
53. Hove, M. J., Marie, C., Bruce, I. C. & Trainor, L. J. Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10383–10388 (2014).
54. Varlet, M., Williams, R. & Keller, P. E. Effects of pitch and tempo of auditory rhythms on spontaneous movement entrainment and stabilisation. *Psychol. Res.* **84**, 568–584 (2020).
55. Sevgi, M., Diaconescu, A. O., Henco, L., Tittgemeyer, M. & Schilbach, L. Social Bayes: Using Bayesian modeling to study autistic trait-related differences in social cognition. *Biol. Psychiatry* **87**, 185–193 (2020).
56. Comstock, D. C., Hove, M. J. & Balasubramaniam, R. Sensorimotor synchronization with auditory and visual modalities: Behavioral and neural differences. *Front. Comput. Neurosci.* **12**, 53 (2018).
57. Eg, R. & Behne, D. M. Perceived synchrony for realistic and dynamic audiovisual events. *Front. Psychol.* **6**, 736 (2015).
58. Hove, M. J., Spivey, M. J. & Krumhansl, C. L. Compatibility of motion facilitates visuomotor synchronization. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 1525–1534 (2010).
59. Varlet, M., Marin, L., Issartel, J., Schmidt, R. C. & Bardy, B. G. Continuity of visual and auditory rhythms influences sensorimotor coordination. *PLoS ONE* **7**, e44082 (2012).
60. Burr, D., Banks, M. S. & Morrone, M. C. Auditory dominance over vision in the perception of interval duration. *Exp. Brain Res.* **198**, 49–57 (2009).
61. Barnhart, W. R., Rivera, S. & Robinson, C. W. Different patterns of modality dominance across development. *Acta Psychol. (Amst)* **182**, 154–165 (2018).
62. Hove, M. J., Iversen, J. R., Zhang, A. & Repp, B. H. Synchronization with competing visual and auditory rhythms: Bouncing ball meets metronome. *Psychol. Res.* **77**, 388–398 (2013).
63. Tryfon, A. *et al.* Auditory-motor rhythm synchronization in children with autism spectrum disorder. *Res. Autism Spectr. Disord.* **35**, 51–61 (2017).
64. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
65. Cheng, M., Kato, M. & Tseng, C. Gender and autistic traits modulate implicit motor synchrony. *PLoS ONE* **12**, e0184083 (2017).
66. Bridges, D., Pittot, A., MacAskill, M. R. & Peirce, J. W. The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ* **8**, e9414 (2020).
67. Brainard, D. H. *The Psychophysics Toolbox Short Title: The Psychophysics Toolbox Corresponding Author.* <http://color.psych.ucsb.edu/psychtoolbox>.
68. Maduell, M. & Wing, A. M. The dynamics of ensemble: The case for flamenco. *Psychol. Music* **35**, 591–627 (2007).
69. Honisch, J. J., Elliott, M. T., Jacoby, N. & Wing, A. M. Cue properties change timing strategies in group movement synchronisation. *Sci. Rep.* **6**, 1–11 (2016).
70. Timmers, R., Endo, S., Bradbury, A. & Wing, A. M. Synchronization and leadership in string quartet performance: A case study of auditory and visual cues. *Front. Psychol.* **5**, 645 (2014).
71. Marieke Van Der Steen, M. C., Jacoby, N., Fairhurst, M. T. & Keller, P. E. Sensorimotor synchronization with tempo-changing auditory sequences: Modeling temporal adaptation and anticipation. *Brain Res.* <https://doi.org/10.1016/j.brainres.2015.01.053> (2015).
72. Repp, B. H. & Keller, P. E. Sensorimotor synchronization with adaptively timed sequences. *Hum. Mov. Sci.* **27**, 423–456 (2008).
73. Colley, I. D., Keller, P. E. & Halpern, A. R. Working memory and auditory imagery predict sensorimotor synchronisation with expressively timed music. *Q. J. Exp. Psychol.* **71**, 1781–1796 (2018).
74. Bishop, L. & Goebel, W. When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance. *Music. Sci.* **19**, 84–110 (2015).
75. Rose, D., Delevoe-Turrell, Y., Ott, L., Annett, L. E. & Lovatt, P. J. Music and metronomes differentially impact motor timing in people with and without Parkinson's disease: Effects of slow, medium, and fast tempi on entrainment and synchronization performances in finger tapping, toe tapping, and stepping on the spot tasks. *Parkinsons. Dis.* **2019**, 6530838 (2019).
76. Lausberg, H. & Sloetjes, H. The revised NEUROGES-ELAN system: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behav. Res. Methods* **48**, 973–993 (2016).
77. Bello, J. P. *et al.* A tutorial on onset detection in music signals. *IEEE Trans. Speech Audio Process.* **13**, 1035–1046 (2005).
78. The jamovi project (2021). jamovi (Version 1.6) [Computer Software]. Retrieved from <https://www.jamovi.org>.

Acknowledgements

We would like to thank Nadyne Dunkley, Makena Peart, Hristina Mihaylova, and Dr Anthony Haffey for helping with the data collection and stimuli creation, Dagmar S. Fraser for his programming support for the stimuli presentation, and thank all the volunteers for participation and feedback. BC was supported by UK Medical Research Council (MR/P023894/1, MR/S036423/1) during this project.

Author contributions

J.J.H., B.C. and O.G. designed the experiment, J.J.H. and B.C. conducted the study, J.J.H., P.M. and B.C. analysed the data and all authors discussed the results. J.J.H., B.C. wrote the paper, and all authors reviewed the 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/s41598-021-88112-y>.

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

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