1

2

3

4

5

6 **Pose estimates from online videos show that side-by-side walkers synchronize movement**

7 **under naturalistic conditions**

8

9 **Synchronization of walking "in the wild"**

10

11

12

13

14 Claire Chambers [1][¶][*], Gaiqing Kong [2][¶], Kunlin Wei [2], Konrad Kording [1]

15

16

17

18 [1] Department of Bioengineering and Department of Neuroscience, University of

19 Pennsylvania, Philadelphia, PA, USA

20

21 [2] School of Psychological and Cognitive Sciences, Peking University, Beijing, China

22

23

24

25 * Corresponding author

26 E-mail: clairenc@seas.upenn.edu

27

28

29 ¶ These authors contributed equally to this work.

30
31
32
33
34
35

Abstract

Marker-less video-based tracking promises to allow us to do movement science on existing video databases. We revisited the old question of how people synchronize their walking using real world data. We thus applied pose estimation to 348 video segments extracted from YouTube videos of people walking in cities. As in previous, more constrained, research, we find a tendency for pairs of people to walk in phase or in anti-phase with each other. Large video databases, along with pose-tracking algorithms, promise answers to many movement questions without experimentally acquiring new data.

Author summary

Lab experiments have shown that people align their movements when they walk together. We asked if walkers synchronize their movements in real-life settings too. To do so, we examined how people move in YouTube videos using marker-less tracking. We confirmed previous results: people walking together tend to walk either in phase or in opposite phase. Marker-less tracking from video allows us to address questions about naturalistic movement using existing video data.

Introduction

To successfully navigate any environment, a walker must adapt to the surface they walk on, avoid obstacles, change speed, and plan movements according to their goals (1,2). Processes that impact walking behavior range from peripheral processes to high-level decision-making processes (3,4). Walking, like all movement, often takes place in the context of other people. Humans, when walking, must often generate their actions according to the movements of people around them.

60   Coordination of movements with others during everyday activities helps to achieve

61   shared goals efficiently and fluently, for example when playing sports, moving an object

62   together, or walking side-by-side. In many cases, synchronization is needed to successfully

63   carry out the task at hand. However, evidence from laboratory experiments suggests that

64   synchronization between people does not only take place when it is required by the task.

65   There is evidence for uninstructed movement coordination from laboratory experiments

66   where subjects are asked to perform various artificial tasks, for example, sway pendulums, sit

67   in rocking chairs, or spontaneously move their arms (5–11). Coordination across individuals

68   thus appears to be fairly ubiquitous, occurring even when it is not required by the task and

69   there is no instruction to do so.

70   Previous laboratory-based studies have investigated if people walking side-by-side

71   synchronize their movements. Walker synchronization is common when people walk side-

72   by-side, and has been found in over-ground walking studies (4,12–14) and in treadmill

73   walking studies (15–20). For example, Zivotofsky and Hausdorff (2007) found that when

74   people were asked to walk side-by-side, they walked in anti-phase (left-right, right-left), with

75   the left leg of one walker aligned with the right leg of the other walker, without instruction.

76   van Ulzen et al. (2008) tracked subjects during side-by-side treadmill walking and found both

77   in-phase (left-left, right-right) and anti-phase (left-right, right-left) walking with preferred

78   synchronization modes. In sum, previous results provide evidence that pairs tend to walk in

79   phase or in anti-phase.

80   However, as with many movement science paradigms, it is not known if walker

81   synchronization occurs in real world settings, outside of the laboratory. This is especially

82   relevant for the walking studies in the laboratory: researchers have found a substantial

83   difference in biomechanics between walking on the treadmill and walking on level ground

84   (21,22). The synchronization of pairs of walkers observed on treadmill might not generalize

85   to real life settings.

86   There have been previous examinations of walker synchronization in naturalistic

87   settings. While most lab studies used motion capture systems to quantify walking parameters,

88   research has relied on reviewer ratings to address walking over ground (14). This is

89   understandable since motion capture technologies were either low in precision, like

90   smartphone sensors, or low in portability, with previous experiments requiring that subjects

91   wear markers or recording equipment (23,24). This made motion capture a challenge for

92   studying walker synchronization in real life. Thanks to recent work on pose tracking from 2D

93   video, it is now possible to perform marker-less tracking with only 2D video data as input

94   (25–27). These algorithms take as input a 2D RGB image and output an estimate of human

95   pose in image coordinates with reasonable accuracy. Marker-less tracking applied to videos

96   provides the opportunity to ask questions about naturalistic behavior outside the laboratory.

97   Here we apply pose estimation to online videos to ask how people synchronize their

98   movements when they walk side-by-side. We used a state-of-the-art pose-estimation

99   algorithm, OpenPose, to extract human body pose from 2D videos (25). OpenPose finds joint

100   positions of people in videos and fits a 2D skeleton model to each person in the image. We

101   built a simple search algorithm to find video segments containing pairs of people walking

102   side-by-side in YouTube videos. From the resulting video segments, we extracted the pose of

103   people in the video and tracked their identity through videos. We examined pose estimates to

104   ask if people walking together synchronized their movements, using vertical ankle

105   displacement to quantify walking angle and thus phase.

106

107   **Methods**

108   **Video data**

109   We wanted to obtain videos of people walking in a naturalistic setting. To find this

110   data, we searched for videos on youtube.com using the search term "walking in" followed by

111   the names of major cities: Bangkok, London, Paris, Dubai, New York, Singapore, Kuala

112   Lumpur, Istanbul, Tokyo, Seoul, Hong Kong, Barcelona, Amsterdam, Milan, Taipei, Rome,

113   Osaka, Vienna, Shanghai, and Prague. Videos were ranked by relevance by YouTube. We

114   examined the most relevant videos for each search term that were over 20 minutes long. In

115   our analysis, we included videos that contained pairs of people walking with low camera

116   movement and where limbs of people in the video were uncropped. We selected videos that

117   met these criteria from the 20 most relevant results on YouTube.

118   **Pose estimation**

119   We needed a method to quantify the synchronization of walker pairs from 2D videos.

120   We therefore began by extracting human pose from video frames, using OpenPose, a pose-

121   estimation algorithm that detects body part positions from 2D RGB images (25). OpenPose

122   detects the position of the neck, shoulders, elbows, wrists, hips, knees, and ankles, as well as

123   key facial points of eyes, ears, and nose. OpenPose assigns a unique person identity to each

124   skeleton present in the image. In our analysis pipeline, we extract the pose of humans from

125   individual frames of videos.

126   **Search algorithm**

127   We first needed to identify video segments of people walking side-by-side from the

128   downloaded YouTube videos. We therefore built an algorithm to search through YouTube

129   videos and extract the particular segments containing pairs of people walking (Fig 1). First,

130   we subsampled YouTube videos at a rate of 2 frames/second, then extracted pose estimates

131   from video frames. Based on pose estimates present in a frame we detected pairs of walkers.

132   Finally, we defined continuous sequences of frames that contained pairs of walkers for

133   inclusion in our analysis.

134   From each frame of the subsampled YouTube video, we extracted human pose,

135   resulting in a set of x, y coordinates indicating body part position and a person index. We

136   included skeletons that contained at least 14 out of the total 18 parts. As movement

137   synchronization is typically quantified by relative phase between two oscillatory motions, we

138   created a measure of walking phase by using the ankle position. Therefore, we required that

139   both ankles be detected for a given skeleton. This resulted in a set of pose estimates for each

140   video frame with people present.

141   Using pose estimates from each frame, we then detected pairs of walkers on a frame-

142   by-frame basis. We built a bounding box around each skeleton pair found in the image. To

143   ensure that walkers were side-by-side, we selected pairs where the aspect ratio (height:width)

144   of the bounding box was between 1:1 and 3:1, and where ankles of the pair were within 15%

145   of the bounding box from each other in y-coordinates. To avoid data with low spatial

146   resolution, we selected pairs where the height of the bounding box was over 120 pixels. To

147   ensure that we analyzed pairs of adult walkers and not adult-child pairs, we only included

148   pairs where the ratio between the larger and smaller skeleton was less than 1.25:1. This

149   resulted in a set of detected person pairs from each video frame.

150   To produce video segments of pairs of walkers, we identified sequences of video

151   frames that contained at least one pair of walkers for the whole duration of the segment, with

152   a minimum segment lasting 2.5 seconds. We excluded pairs where two or more walker pair

153   detections contained a common element (person) for 3 consecutive subsampled frames or

154   more (1.5 seconds), which would suggest a group of walkers. We then extracted video

155   segments of pair of walkers from the original YouTube video sampled at the original frame

156   rate. We manually selected videos that contained non-occluded pairs of adults walking

157   continuously toward or away from the camera on flat surfaces in well-lit conditions, where

158   people were not carrying large objects or using an assistive device. This resulted in video

159   segments containing pairs of people walking together.

160   From each video segment, we needed to extract pose and track individuals through the

161   video, in order to obtain cyclic walking signals for each person. We extracted pose estimates

162   for all frames. Using the pose estimates, we tracked pairs of walkers through the frames of

163   the video. We tracked pairs of walkers by first running the previous walker pair detection step

164   on the first frame of the segment. We then minimized the distance between skeletons in the

165   video from one frame to the next averaged across the key points present in both frames, $D =$

166      $\sum_i^n \frac{1}{n}\sqrt{(x_{i,j} - x_{i,j+1})^2 + (y_{i,j} - y_{i,j+1})^2}$ where i is a keypoint index, j is a frame, n is the

167   number of keypoints that are present in both frames, x is the x-coordinate of a key point and

168   y is the y-coordinate of the key point. To track a given person, we iterated through the frames

169   of the video and minimized the distance for each frame pair. Common problems were

170   switches in person identity across frames and failure to accurately track the ankle. We

171   therefore manually selected videos for inclusion in our analysis where the algorithm

172   successfully tracked people through frames of the video.
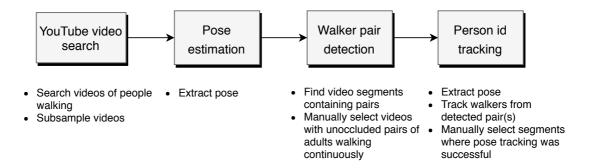
YouTube video search	→	Pose estimation	→	Walker pair detection	→	Person id tracking
• Search videos of people walking • Subsample videos		• Extract pose		• Find video segments containing pairs • Manually select videos with unoccluded pairs of adults walking continuously		• Extract pose • Track walkers from detected pair(s) • Manually select segments where pose tracking was successful

173

174 **Fig 1.** Flowchart describing stages of search algorithm used to find segments of videos
175 containing pairs of walkers.
176

177 **Data analysis**

178   To quantify walker synchronization from pose estimates, we analyzed the vertical

179   displacement between the ankle joints, using this as a way of quantifying the relative walking

180   phase. We first replaced missing values using linear interpolation. We did not have control

181   over video recordings and so, in most video segments, the distance between the camera and

182   the walkers was not fixed. We therefore needed to detrend the ankle coordinates. To do so,

183   within each frame, we computed the distance of the ankle relative to a reference body part

184   and divided this distance by the height of the skeleton. We selected the neck as the reference

185   part since this is typically the least occluded part. We combined this data across frames to

186   form a time series of detrended ankle coordinates. We removed outliers from the signal by

187   removing points which were two standard deviations or more from the mean. We then

188   normalized the signal by subtracting its mean and dividing by its standard deviation. We low-

189   pass filtered the ankle displacement signal using a third order Butterworth filter with a cutoff

190   frequency of 2 Hz. This resulted in a signal from which we could extract the vertical ankle

191   displacement and walking phase.

192   For each pair of walkers, we wanted to analyze the walking phase relationship using

193   the positions of the ankles. We computed the relative ankle displacement by subtracting the

194   right ankle position from the left ankle position. We then computed the Hilbert phase of the

195   displacement signal. The phase signal, which we use as a proxy for oscillatory walking

196   motion for one person, varied between $-\pi$ and $+\pi$. The relative phase between walkers was

197   simply computed by subtracting the walking phase of one walker from the other. We also

198   wrapped the phase difference so that it could not extend outside $-\pi$ and $+\pi$. This resulted in

199   a time series of relative phases for each pair of walkers. We analyzed the distribution of

200   relative phase for each pair of walkers by computing the mean phase angle.

201   In order to combine estimates of walker synchronization across different videos, we

202   computed the mean of the phase distribution from each video segment. We fitted mixtures of

203   von Mises distributions to the distribution of mean relative phase extracted from each video,

204   and identified the peaks of the best fitting distribution. Peak phase of 0 would mean that

205   walking in phase is a dominant mode of walker synchronization, and peak phase of π would

206   mean that walking in anti-phase is a dominant mode.

207   We extracted the frequency of walking by examining displacement between left and

208   right ankles in y coordinates. The displacement signal is positive when the left ankle is above

209   the right and negative when the left ankle is below the right. This displacement signal is

210   cyclic, and one cycle represents two steps. We extracted the frequency by computing the peak

211   value of the discrete fast Fourier transform. We doubled this frequency in cycles/second to

212   obtain the walking frequency conventionally defined as steps/second.

213   In our video data, we had no control over how the video was recorded. For example,

214   the camera was often in movement and camera angle was variable, both of which could have

215   an influence on our measure of ankle position. We therefore wanted to ensure that our

216   analyses of frequency and relative walking phase were unbiased. We verified that the mean

217   relative phase extracted from pose estimates was consistent with the mean relative phase

218   extracted from labels of foot strike timings. Our "ground-truth" data set consisted of 5-

219   second-long segments from 43 randomly-selected videos from our final data set. We

220   annotated left and right foot strike timings for each member of the walking pair. Note that

221   foot strike is one of the most salient kinematic event for locomotion, and its timing is easy to

222   estimate (28,29). We assumed that the maximal displacement between left and right ankles

223   took place at the foot strike. From the foot strike timings, we built a piecemeal sine wave

224   where +1 represented left foot strike and -1 represented a right foot strike. We compared the

225   frequency and peak relative phase extracted from this signal with that extracted from pose

226   estimates. This allowed us to validate our analysis from pose estimates.

227   We wanted to ensure that the ages of pairs found in videos were not biased to a

228   certain age range, e.g., youth. For the videos in our final data set, we estimated the ages of

229   walkers in each pair to the nearest decade.

230   This study was approved by the University of Pennsylvania Institutional Review

231   Board (protocol 827723).

232

233   **Results**

234   We asked if people synchronize their walking when they walk side-by-side in

235   naturalistic settings. To do so, we analyzed videos found on YouTube. Within YouTube

236   videos, we searched for video segments with pairs of people walking. From video segments,

237   we extracted the pose of each member of the walking pair. In order to examine walker

238   synchronization, we analyzed displacement between left and right ankles for each member of

239   the pair. Based on the displacement signal, we extracted the walking frequency and mean

240   relative phase for each pair to examine walker synchronization.

241   To find relevant videos, we searched for videos on youtube.com using the search term

242   'walking in' followed by the names of major cities. Our search resulted in 363 videos. We

243   excluded 113 videos because they did not include footage of pairs of people walking. We

244   excluded 48 videos because of large amounts of camera movement, cropped video frames,

245   camera angles which prevented pose estimation, and poor visibility. We excluded 3 videos

246   because the format of the video was not suitable (a large amount of occluding text on the

247   screen and panoramic videos). After screening, we included a total of 199 videos, which gave

248   us a reasonably sized dataset for quantifying synchronization.

249   Within these videos, our algorithm identified 888 video segments with pairs of people

250   walking lasting at least 2.5 seconds. Some segments were unsuitable because 1.) the segment

251   did not contain pairs of adults walking continuously towards or away from the camera in

252   well-lit conditions (258 segments), 2.) the segment included large amounts of camera

253   movement, occlusion, or cropping of body parts from the video which prevented successful

254   pose-tracking (197 segments). In the end, we extracted pose from the 433 suitable video

255   segments containing 441 pairs of walkers, among which pose estimation or tracking failed for

256   93 walking pairs. Thus, our final sample for walking analyses include 348 video segments

257   with a mean duration of 4.73 seconds (SD = 5.99, Fig 2). Walkers in the video had a mean

258   age rating of 30.66 years (SD = 12.34). Although most walkers were in the young age range

259   of 20~40 years (71%), we still had 29% of walkers in the range of 40 to 70 years. Thus, we

260   have a reasonably representative dataset to investigate our questions.
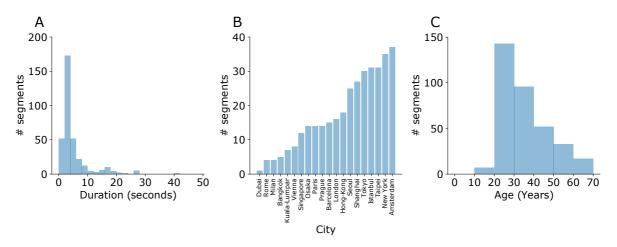
261
262   **Fig 2.** Descriptive data of videos. (**A**) Distribution of duration of video segments in seconds.
263   (**B**) Number of video segments for each city we searched for. (**C**) Distribution of estimated
264   age of walkers averaged across the pair in years.
265

266   We wanted to quantify the degree of synchronization in the walking phase of each

267   pair. To do so, we examined the phase relationship between two oscillatory walking patterns,

268   using the signed y-displacement between left and right ankles. From the time series of left

269   and right ankle positions in y-coordinates (Fig 3A, B), we extracted the cyclic motion of

270   walking for each member of a walking pair. We first detrended the y-coordinates of the left

271   and right ankles (Fig 3C, D). Then we normalized and low pass filtered the detrended signal

272   (Fig 3E, F). For each walker, we computed the displacement between the left and right ankles

273   (Fig 3G), then computed the Hilbert phase of the displacement time series (Fig 3H). Finally,

274   we computed the relative phase between Walker 1 and Walker 2. We examined the

275   distribution of the relative phase to quantify walker synchronization (Fig 3I). The

276 representative video segment in the Fig 3 showed a predominant in-phase synchronization as

277 the relative phase clustered around zero. The ability to estimate relative phase from videos

278 enables us to estimate synchronization behavior in the population.
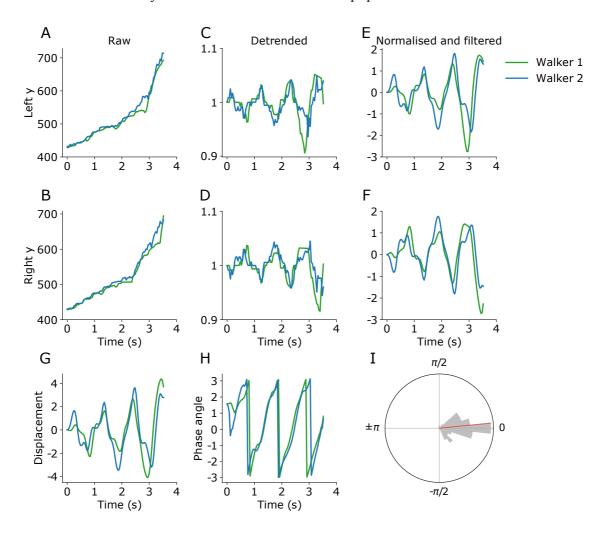

279

Fig 3. Extracting relative phase from the y-coordinates of the left and right ankles. Walkers
are superimposed on the same panel. The strong overlap between the signals shows that the
pair is walking in phase. (**A, B**) We first extracted the y-coordinates of the left and right
ankles for Walker 1 and Walker 2 frame by frame. (**C, D**) We then detrended the signal by
subtracting the y-coordinate of a reference body part, the neck, and dividing by the height of
the walker. This gave a signal that was invariant to the position of the walkers in the frame
and their size. (**E, F**) We then removed outliers, normalized and low-pass filtered the signal.
(**G**) Using the normalized y-coordinate signal, we subtracted right ankle coordinate from the
left ankle coordinate. (**H**) We computed the phase angle by taking the Hilbert transform of
the difference. (**I**) We examined the distribution of the relative phase between walkers by
computing the mean of the distribution (red).

291

292   We wanted to validate the outputs from pose estimates. We thus compared the

293 walking frequency and mean relative phase computed from pose estimates with those

294 extracted from a ground truth signal. We found reasonable agreement between the ground-

295 truth estimates and pose estimates (Fig 4, relative phase: $R^2 = .71$ (N = 43); walker frequency:

296 $R^2 = .23$ (N = 43)). Therefore, despite the variability of recording conditions, we were able to

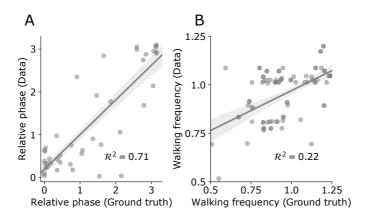297 meaningfully extract information about walking patterns from pose estimates.

298

299 **Fig 4.** Comparison of relative phase and walking frequency computed from pose estimates
300 and ground truth. (**A**) Mean relative phase from pose estimates as a function of mean relative
301 phase from ground truth data with overlaid linear fit. (**B**) Walker frequency from pose
302 estimates as a function of walker frequency from ground truth data with overlaid linear fit.
303

304   We asked what the overall synchronization is in our sample as a whole. For an

305 analysis that combined data across video segments, we quantified the mean relative walking

306 phase for each video segment and examined its distribution across all segments. To quantify

307 the most prevalent relative phases, we fit a mixture of von Mises distributions model to the

308 distribution of mean relative phase, comparing models with one to four peaks. We found that

309 the model with three peaks provided the best fit to the data, as quantified by Bayesian

310 Information Criterion (Fig 5A). The means of the three components or peaks of the model

311 were .01, -3.05, and -3.13 in radians with concentrations (normalized κ) of .17, .80, and .02

312 respectively (Fig 5B). The component with approximately zero mean shows that pairs walked

313 in phase with each other. The other peak close to π show that pairs also walk in near anti-

314   phase with each other. Note, the two peaks near π are close to each other and, accordingly,

315   the two-peak model is comparable to the three-peak model in BIC. We have thus shown that

316   people walking in pairs have a tendency to walk in phase or in anti-phase with each other.
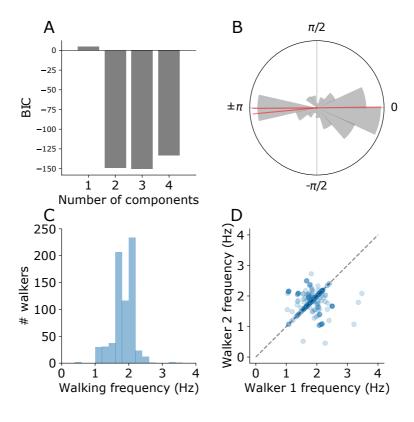

318   **Fig 5.** Group results of relative phase and walking frequency. (**A**) We fitted mixture of von
319   Mises distribution models to the distribution of mean relative phase, varying the number of
320   peaks of the distribution. The best fitting model contains three peaks. (**B**) The distribution of
321   the mean relative phase across pairs of walkers and means of the best-fitting mixture of von
322   Mises distributions (red). (**C**) Distribution of walking frequency, quantified using ankle
323   displacement (Mean = 1.85 Hz, SD = .32). (**D**) Scatterplot of walking frequencies of walkers
324   within each pair.

326   As an additional validity test of our method, we tested if its walking frequency

327   estimate agrees with the typical walking cadence in the literature. We thus measured walking

328   frequency from ankle displacement signals (Fig 5C). The average walker frequency was 1.85

329   Hz (SD = .32), which is very similar to the average of ~2 Hz previously reported in the

330   literature (30). The range of walking frequency was approximately from 1 to 2.5 Hz,

331   appearing to reflect the cadence in naturalistic settings. In fact, the frequency within pairs was

332 almost identical for some walking pairs, shown by points close to the diagonal line in Fig 5D.

333 The frequency of walkers differed by less than .1 Hz in 66% of walking pairs. This was not

334 the case for all walking pairs, though this is understandable as the walking pairs differed

335 (e.g., heights) and our pose estimation is inevitably plagued by noise.

336

337 **Discussion**

338   We asked how people synchronize their movements when they walk side-by-side, by

339 analyzing pose estimates extracted from online videos. We searched for videos on

340 youtube.com using the search term 'walking in' followed by the names of major cities. We

341 analyzed the relative phase and walking frequency computed from the vertical displacement

342 between the ankles in each video. We validated our analysis from pose estimates through

343 comparison with ground-truth data and found reasonable agreement between estimates of

344 mean relative phase and walking frequency. The mean frequency of walkers was 1.85 Hz,

345 close to the average of 2 Hz reported in the literature (30). We found that the distribution of

346 relative phase across video segments contained prominent peaks near 0 and π, which reflects

347 a tendency for pairs of walkers to walk in phase or in anti-phase with each other. Using pose

348 estimation applied to online videos, we confirmed findings from more constrained laboratory

349 experiments on a larger sample and in more naturalistic conditions. We have thus shown that

350 a pose estimation algorithm applied to 2D videos can be used to address questions about

351 naturalistic movement.

352   The strong tendency for walking in-phase or anti-phase is consistent with previous

353 laboratory studies (4,12,17). However, previous studies have to examine walking pairs on a

354 treadmill, which is set at a fixed walking speed (17). Furthermore, biomechanical studies

355 have found that normal ground walking is different from treadmill walking (21,22). The

356 manipulation of walking speed and the very fact of walking on the treadmill might impact the

357   synchronization results and prevent generalization of previous findings. Our study bypassed

358   these questions by investigating unconstrained, naturalistic walking from videos shot in

359   different cities. In these videos, people walk with their "preferred" speed, and their leg

360   lengths presumably vary widely. Despite the lack of experimental control, we still find

361   evident phase entrainment between walking patterns. Hence, our study provides solid

362   evidence for unintentional synchronization between walking pairs.

363   We developed a pipeline to study walking behavior from videos. Using this pipeline,

364   we were able to extract sufficient information to carry out a group analysis across hundreds

365   of video segments. Outside the scope of this work, various interesting questions arise on the

366   dynamics between individual walkers. For example, walkers in social situations created in the

367   laboratory have been found to obey leader-follower dynamics (2) and while pairs spend more

368   time walking in phase or in anti-phase, the phase relationship has been found to shift over

369   time (17). Another interesting question to ask is whether biomechanical features such as leg

370   length impact walking synchronization, as shown in the laboratory (18). Addressing these

371   questions would require a more detailed analysis of individual data, precluded by the level of

372   noise in our data and the lack of longer video segments in our dataset. Larger data sets and

373   more precise pose estimation algorithms will make these analyses possible in future.

374   Although pose estimation is a promising tool for movement science and related

375   applications, its application at present involves limitations. One challenge is the presence of

376   different sources of variability in videos. Due to imperfect pose estimation, our measure of

377   ankle joint position varied around the true position. To remove noise, more sophisticated

378   forms of denoising than the standard signal processing techniques used here can be applied.

379   For example, measuring the contribution of different noise sources would allow the

380   application of deconvolutional techniques to separate the walking phase signal from noise.

381   Another problem is that camera angle, which constantly changes in the videos that we used,

382 influenced the ankle displacement variable. Signal cleaning that uses machine-learning based

383 predictions from, for example, recurrent neural networks would allow the algorithm to learn

384 the mapping of 2D-joint positions to walking phase (31). Although standard techniques were

385 sufficient for our purposes here, we anticipate the need to apply more sophisticated

386 techniques to address questions that require more precise measurements when using real

387 world data.

388 A second limitation is that most existing pose-estimation algorithms output two-

389 dimensional pose estimates (25,26). Without depth cameras, it is therefore impossible to

390 measure true three-dimensional kinematic variables. For example, in the present study, this

391 prevented us from measuring walking phase directly from joint angles. Answering questions

392 about naturalistic movement and developing usable applications generally requires estimation

393 of a subjects' three-dimensional pose. Some existing algorithms that fit a three-dimensional

394 body model to two-dimensional images may provide a solution (27,32). However, they are,

395 for the most part, inaccessible to those without expertise in the field of computer vision. The

396 release of these algorithms in the form of more accessible software will contribute to the

397 advancement of movement science and many fields in behavioral science.

398 A third challenge is continuous tracking of person identity in videos. Pose estimation

399 algorithms provide joint position coordinates for each frame but do not track person identity

400 throughout a video. Simple tracking algorithms that minimize distance across frames often

401 fail because large amounts of occlusion present in natural scenes. Improvements of

402 optimization methods for tracking algorithms and integrating pose estimation with tracking

403 on whole videos will allow more successful analysis of movement in natural and dynamic

404 scenes (33,34).

405 Besides algorithmic solutions, we can devise practical solutions using specialized

406 setups to meet the above challenges. For example, we could carefully place multiple cameras

407   in naturalistic settings. Alternatively, we could place lab-grade sensors on people in the real

408   world. While all these approaches are doable and have been used in past research, having

409   scalable approaches that can leverage large datasets is particularly promising.

410   Many real-world applications are calling for automatic pose estimation based on

411   videos. Improved pose tracking will allow the field to quantify movement variables from

412   easily obtained data, thus paving the way for a host of video-based applications in fields

413   including medicine and sport (35). Movement tracking applications will allow us to better

414   quantify the parameters of movement indicative of disease and thus will enable us to build

415   objective, inexpensive, and affordable video-based diagnostic software that can replace

416   expensive in-person evaluations. Similarly, pose tracking can be used to quantify the success

417   of physical therapy interventions and will provide an objective standard that will help to

418   guide therapists. Applications for sports, based on the parameterization of the best possible

419   movement and other relevant variables related to fatigue or injury, include "virtual coaching"

420   where an application can instruct users on when to train or how to best execute a movement

421   from video recordings. We expect that such applications will soon be integrated into standard

422   practice within these fields.

423   Movement science stands to gain the most from adopting pose estimation as part of its

424   standard toolkit. Movement science and the field of psychology more generally have been

425   limited by small data set sizes and a set of overly constrained laboratory tasks that may

426   invoke qualitatively different mechanisms to those at play during real world tasks (36,37).

427   Examination of 3D kinematics of real-world movement data will capture important aspects of

428   behavior unseen during simple laboratory tasks. Such an examination alone may be sufficient

429   to inform models. Pose estimation can be used in combination with an experimental approach

430   to constrain the kinds of experiments we do in the laboratory, and to provide evidence for or

431   against models based on real-world data. Using videos will make large amounts of kinematic

432   data easy to collect and will remove the need to collect new data in some cases, since existing

433   video data sets can be used. To progress, movement science must embrace a more ecological

434   approach that emphasizes movement as it occurs "in the wild" (35,37) and we expect that

435   pose estimation from video will be key to this process.

436

437   **References**

438   1.   Seethapathi N, Srinivasan M. The metabolic cost of changing walking speeds is

439   significant, implies lower optimal speeds for shorter distances, and increases daily

440   energy estimates. Biol Lett. 2015;11(9):20150486.

441   2.   Ducourant T, Vieilledent S, Kerlirzin Y, Berthoz A. Timing and distance

442   characteristics of interpersonal coordination during locomotion. Neurosci Lett.

443   2005;389(1):6–11.

444   3.   Borghese NA, Bianchi L, Lacquaniti F. Kinematic determinants of human locomotion.

445   J Physiol. 1996;494(3):863–79.

446   4.   Zivotofsky A, Bernad-Elazari H, Grossman P, Hausdorff J. The effects of dual tasking

447   on gait synchronization during over-ground side-by-side walking. Hum Mov Sci.

448   2018;59:20–9.

449   5.   Issartel J, Marin L, Cadopi M. Unintended interpersonal co-ordination: "can we march

450   to the beat of our own drum?" Neurosci Lett. 2007;411(3):174–9.

451   6.   Demos AP, Chaffin R, Begosh KT, Daniels JR, Marsh KL. Rocking to the beat:

452   Effects of music and partner's movements on spontaneous interpersonal coordination.

453   J Exp Psychol Gen. 2012;141(1):49–53.

454   7.   Miles LK, Griffiths JL, Richardson MJ, MacCrea N. Too late to coordinate:

455   Contextual influences on behavioral synchrony. Eur J Soc Psychol. 2010;40:625–34.

456   8.   O'Brien B, Schmidt RC. Evaluating the Dynamics of Unintended Interpersonal

457   Coordination. Ecol Psychol. 2010;7413:37–41.

458   9.   Richardson MJ, Marsh KL, Schmidt RC. Effects of visual and verbal interaction on

459   unintentional interpersonal coordination. J Exp Psychol Hum Percept Perform.

460   2005;31(1):62–79.

461 10.   Richardson MJ, Marsh KL, Isenhower RW, Goodman JRL, Schmidt RC. Rocking

462   together: Dynamics of intentional and unintentional interpersonal coordination. Hum

463   Mov Sci. 2007;26(6):867–91.

464 11.   Schmidt RC, Richardson MJ. Dynamics of interpersonal coordination. Underst

465   Complex Syst. 2008;281–308.

466 12.   Sylos-Labini F, D'Avella A, Lacquaniti F, Ivanenko Y. Human-Human interaction

467   forces and interlimb coordination during side-by-side walking with hand contact. Front

468   Physiol. 2018;9(179):1–14.

469 13.   Zivotofsky A, Gruendlinger L, Hausdorff JM. Modality-specific communication

470   enabling gait synchronization during over-ground side-by-side walking. Hum Mov Sci.

471   2012;31(5):1268–85.

472 14.   Zivotofsky A, Hausdorff JM. The sensory feedback mechanisms enabling couples to

473   walk synchronously: An initial investigation. J Neuroeng Rehabil. 2007;4:1–5.

474 15.   Nessler JA, Gutierrez V, Werner J, Punsalan A. Side-by-side treadmill walking

475   reduces gait asymmetry induced by unilateral ankle weight. Hum Mov Sci.

476   2015;41:32–45.

477 16.   Nessler JA, Kephart G, Cowell J, De Leone CJ. Varying treadmill speed and

478   inclination affects spontaneous synchronization when two individuals walk side by

479   side. J Appl Biomech. 2011;27(4):322–9.

480 17.   van Ulzen NR, Lamoth CJC, Daffertshofer A, Semin GR, Beek PJ. Characteristics of

481   instructed and uninstructed interpersonal coordination while walking side-by-side.

482 Neurosci Lett. 2008;432(2):88–93.

483 18. Nessler JA, Gilliland SJ. Interpersonal synchronization during side by side treadmill

484 walking is influenced by leg length differential and altered sensory feedback. Hum

485 Mov Sci. 2009;28(6):772–85.

486 19. Nessler JA, De Leone CJ, Gilliland S. Nonlinear time series analysis of knee and ankle

487 kinematics during side by side treadmill walking. Chaos An Interdiscip J Nonlinear

488 Sci. 2009;19(2):026104.

489 20. Nessler JA, Gilliland SJ. Kinematic analysis of side-by-side stepping with intentional

490 and unintentional synchronization. Gait Posture. 2010;31(4):527–9.

491 21. Lee SJ, Hidler J. Biomechanics of overground vs. treadmill walking in healthy

492 individuals. J Appl Physiol. 2008;104(3):747–55.

493 22. Ochoa J, Sternad D, Hogan N. Treadmill versus Overground Walking: Different

494 Response to Physical Interaction. J Neurophysiol. 2017;118:2089–2102.

495 23. Hayhoe M, Ballard D. Eye movements in natural behavior. Trends Cogn Sci.

496 2005;9(4):188–94.

497 24. Howard IS, Ingram JN, Kording KP, Wolpert DM. Statistics of Natural Movements

498 Are Reflected in Motor Errors. J Neurophysiol. 2009;102(3):1902–10.

499 25. Cao Z, Simon T, Wei S-E, Sheikh Y. Realtime Multi-Person 2D Pose Estimation using

500 Part Affinity Fields. IEEE Conf Comput Vis Pattern Recognit. 2016;1–9.

501 26. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al.

502 DeepLabCut: markerless pose estimation of user-defined body parts with deep

503 learning. Nat Neurosci. 2018;21(9):1281–9.

504 27. Güler RA, Neverova N, Kokkinos I. DensePose: Dense Human Pose Estimation In

505 The Wild. arXiv. 2018;1802.00434.

506 28. Hreljac A, Marshall RN. Algorithms to determine event timing during normal walking

507   using kinematic data. J Biomech. 2000;33:783–6.

508   29.   Mickelborough J, Van Der Linden ML, Richards J, Ennos AR. Validity and reliability

509   of a kinematic protocol for determining foot contact events. Gait Posture.

510   2000;11(1):32–7.

511   30.   Pachi A, Ji T. Frequency and velocity of people walking. Struct Eng. 2005;83(3):36–

512   40.

513   31.   Maas AL, Le Q V., O'Neil TM, Vinyals O, Nguyen P, Ng AY. Recurrent Neural

514   Networks for Noise Reduction in Robust ASR. In: Thirteenth Annual Conference of

515   the International Speech Communication Association. 2012.

516   32.   Loper M, Mahmood N, Romero J, Pons-moll G, Black MJ. SMPL: A skinned multi-

517   person linear model. In: ACM Transactions on Graphics (TOG). 2015. p. 248.

518   33.   Iqbal U, Milan A, Gall J. PoseTrack: Joint multi-person pose estimation and tracking.

519   IEEE Conf Comput Vis Pattern Recognit. 2017;4654–63.

520   34.   Insafutdinov E, Andriluka M, Pishchulin L, Tang S, Levinkov E, Andres B, et al.

521   ArtTrack: Articulated multi-person tracking in the wild. IEEE Conf Comput Vis

522   Pattern Recognit. 2017;1293–301.

523   35.   Wei K, Kording KP. Behavioral tracking gets real. Nat Neurosci. 2018;21(9):1146–7.

524   36.   Open Science Collaboration. Estimating the reproducibility of psychological science.

525   Science. 2015;349(6251):1–7.

526   37.   Ingram JN, Wolpert DM. Naturalistic approaches to sensorimotor control. Prog Brain

527   Res. 2011;191:3–29.

528