Postprint: This article has been accepted for publication in American Psychologist. The final published version will differ.

An Open-Source Standardized Pipeline for Equitable Observations of Interactive Behavioral Dynamics: Theory-driven Measurement, Analysis, and Masking

Arkadiusz Białek[1#], Wim Pouw[2,3#], James Trujillo[4], Fred Hasselman[5], Babajide Alamu Owoyele[6], Natalia Siekiera[1], Joanna Rączaszek-Leonardi[7] & Travis J. Wiltshire[2]

[1] Institute of Psychology, Jagiellonian University

[2] Department of Computational Cognitive Science, Research Center for Cognitive Science & Artificial Intelligence, Tilburg University

[3] Donders Institute for Brain, Cognition and behavior, Radboud University Nijmegen

[4] Institute for Logic, Language and Computation, University of Amsterdam

[5] Behavioral Science Institute, Radboud University Nijmegen

[6] Artificial Intelligence and Intelligent Systems, Hasso Plattner Institute Potsdam

[7] Human Interactivity and Language Lab, Faculty of Psychology, University of Warsaw

[#]These authors contributed equally

Author Note: The authors declare no financial or other conflicts of interest. Some text may match our extended method and results notebook text in verbatim: https://wimpouw.github.io/InterPerDynPipeline/.

Open Data Statement: See 'Transparency and Openness' statement in the methods.

Public significance statement: This position paper argues for the value of open-source pipelines for collecting and analyzing behavioral data from video recordings. It demonstrates the promise of advances in computer vision to track body movements and applies statistical analysis to identify patterns in behavior. The pipeline includes a workflow for masking recordings, facilitating data archiving and reuse while protecting participant privacy.

Abstract

Behavioral observation has a longstanding tradition in psychology. However, its reliance on manual coding makes observational methods costly and prone to biases that arise from the inherent limits of coding schemes and from the subjectivity of human raters. Behavioral observation research therefore has drawbacks in terms of applicability, accessibility, reproducibility, and equitability. Computer vision (CV) methods have promise in this regard. In the context of quantifying human movement, CV approaches to pose estimation are becoming increasingly precise and have the benefit of being non-intrusive, reproducible, and low-cost. With such methods, human movement in videos can be processed as multivariate time-series data, offering a basis for a more comprehensive analysis of individual, interpersonal, and social dynamics, opening up the possibility for the widespread study of diverse populations in out-of-lab real-world contexts. Fulfilling such a promise goes beyond technical progress, it requires a theory-driven implementation of new accessible observational methods made equitable to learn. To showcase this promise, we present a CV pipeline with analyses of the temporal dynamics of dyadic interactions, including a cross-cultural investigation of movement fluidity in sibling cooperation during task-focused interactions and longitudinal assessments of infant-caregiver coordination during free-play sessions. The pipeline incorporates a pedagogical computational notebook, pose estimation (YOLOv8), theory-driven linear and non-linear time series analysis, and a CV method for more privacy-aware video data archiving. Together, this pipeline showcases a promising advancement in behavioral observation methods, expanding both who can conduct research and who can benefit from it.

Key words: behavioral observation, dynamical systems, interpersonal interaction. open-science, reproducibility

Introduction[1]

When a mother and infant look at each other and exchange smiles or when siblings negotiate who places the next block on a tower, these seemingly simple human interactions reveal rich information about relationships, individual experiences, development, and social functioning. The observation of behavior has a long-standing tradition in the life and social sciences and has been used in science at least since Darwin's work on emotion (Darwin, 1872). Yet capturing and analyzing such behavioral dynamics has remained surprisingly challenging, limiting our theoretical understanding and our ability to study diverse populations equitably.

Observational methods form the core of many studies across disciplines including ethology (Tinbergen, 1963), medicine (Berk & Verghese, 1990), cultural anthropology (Bernard, 2017), and many areas of psychology including educational and clinical psychology. Specifically, observation is a primary mode of gaining knowledge about individual and collective behavior. Observations are realized in a particular time and place, whereby the observer, tuned by evolution and experience, captures functional and socially significant aspects of behavior in specific situations (De Jaegher, 2021). Observational methods often utilize an agreed-upon catalog of behavior codes - a "coding scheme" (Brauner, 2018). Although researchers adopt various strategies to ensure their observations are rigorous, they remain prone to biases, stereotypes, and practical constraints.limits. Biases are inherent to human perception, but also structurally imposed by the scientific community through the historical continuation of drawing upon researchers and participants from WEIRD populations). Additionally, observational methods are extremely time-consuming and therefore limiting. They require highly trained coders and many hours of coding reliability, which makes rigorous observational research expensive and only performable by well-resourced labs. The high potential for bias as well as their resource intensiveness greatly reduces the equitability of these methods.

[1] This article is the result of a collaboration initiated during the Behavioral Dynamics in Social Interactions workshop held at Jagiellonian University on May 2–4, 2024. As such, it reflects our particular approach to the issue of incorporating observation in psychological research. Our approach stems from the ethological tradition, which emphasizes the importance of observation and description of phenomena, as well as from a processual and, in particular, dynamical systems approach to understanding them. Our aim of creating a pipeline for collecting and analyzing observational data reflects the values of equity, transparency, and reproducibility in psychological research.

Advancements in computer vision (CV), i.e. use of computer science techniques to automate the analysis and/or interpretation of images and videos (Chai et al., 2021) may address these limitations. The technique enables motion tracking in more ecological contexts, in a reproducible way, and, being open-source and time-efficient, can be used more widely as an observational tool. Tools such as OpenPose or Mediapipe provide rich multivariate time-series, which can serve as a basis for a less biased, and more comprehensive analysis of individual, interpersonal, and social dynamics across diverse populations. Such rich data and novel analyses enable addressing various new and longstanding research questions and equitability concerns (e.g., Paxton & Dale, 2017). Pose estimation from video may form an important antidote to the generally expensive equipment needed for human movement research such as motion tracking systems which often use software that is not open source. While such high-precision equipment is not yet fully replaceable by CV, it remains inequitable - accessible to only certain labs around the world, and not reusable in a fully transparent way (for discussions on barriers for open science practices see (Bahlai et al., 2019). An additional limitation of high-precision motion tracking systems is the need to attach markers to the body, which some participants, such as children may find intrusive and movement-restricting, thereby reducing the ecological validity and equity of research. CV-based pose estimation finds the position of structurally defined points on the entire body from images or videos (Gao et al., 2025) without need of markers and expensive equipment. These tools further fulfill their promise as observational tools if they are applicable outside of the lab, so as to be integrated into the study of real-world (social) phenomena and enable the investigation of cognition in the wild. Although CV approaches to the study of social interaction are not new (e.g., De Barbosa, 2017), recent developments in these methods have enabled much more precise measurement (e.g., tracking a single body part, such as the wrist, rather than an entire indicated area). In order to fulfill the promise that CV methods hold with respect to becoming more equitable tools for inferring knowledge from movement-based interactions, these tools need to be made accessible and usable for all researchers.

In this paper, we demonstrate the continued and evolving usefulness of CV-based techniques for studying behavior and social interaction in general (for earlier work see Barbosa & Vatikiotis-Bateson, 2006; Latif et al., 2014), and particularly for research grounded in a process-oriented perspective (van Geert & de Ruiter, 2022). To this end, we propose a pipeline that facilitates the observation of naturalistic and semi-naturalistic human interactions. This pipeline integrates pose estimation with analyses based in dynamical systems theory and further combines with masking methods that have promise to increase ethical data preservation/storage. This pipeline is not intended to be a one-size-fits-all solution for all interaction data, but rather a functional workflow going from raw videos to statistical analysis, where specific pieces (e.g., the specific statistical analysis script) can be replaced or adapted to meet researchers' needs.

The pipeline i) relies on video recordings, which are easy to collect and widely available, thereby enabling research in ecologically valid settings and expanding the range of populations that can be readily studied (e.g., individuals who are unable to visit a laboratory due to socioeconomic status, disability, or geographic distance can record themselves in settings that are suitable for them); ii) uses CV analysis of recordings, which provides precise time-series description of behavior based on estimations of the individuals musculoskeletal positions; iii) demonstrates its usability through two case studies; iv) contains a protocol for masking recordings and archiving data; v) is open source and designed for researchers who do not yet have the programming experience to build a pipeline themselves. In sum, the pipeline enables the collection, storage and sharing of precise, multidimensional behavioral data that were previously inaccessible, facilitating a more inclusive theory development (Dale et al., 2023). As such, our approach is also aligned with calls for increased collaboration in revisiting and updating ethical standards for research involving human participants (Light et al., 2024).

Below, we first describe the theoretical underpinnings of the method/pipeline, against the background of traditional observational methods. Importantly, we address the problem that although CV-processed behavioral data may be considered "raw observations" free from coder biases, such data captures movements of individuals and interacting humans that are complex and governed by multiple goals and values, therefore, the individuation from high-dimensional spatio-temporal data to *relevant behaviors* is not straightforward. However, knowledge about the structure of human movement, grounded in dynamical systems theory, facilitates decisions about which variables in a raw stream of data need to be traced and how to analyze and interpret them using theoretical and experiential insights. Against this theoretical background, we describe the pipeline in detail, highlighting its features mentioned above and, finally, illustrate its application in two example cases: the flow of cooperation, operationalized as movement smoothness (Mejía-Arauz et al., 2018) among Yurakaré and Polish siblings (Case 1) and the development of coordination in parent-infant interactions in Poland (Case 2).

Direct Observation of Behavior

Direct observation of behavior is regarded as the oldest research method in behavioral science (Jersild & Meigs, 1939), but due to its labor-intensiveness and its inherent biases, it has never been a substantial mode of inquiry. The human coder is often more than a "detector" of behaviors, serving as "cultural informant" who infers meaning about observed individuals (Bakeman & Gottman, 1997). At the same time, psychology, in its effort to establish itself as mature science, modeled itself on natural sciences and focused on advancing precise and experimental methods (Rozin, 2001). As a result, many branches of psychology have become what Rai and Fiske (Rai & Fiske, 2010) call an ODD science - observation- and description-deprived - relying heavily on self-reports and finger "movements" (Doliński, 2018).

The proposed pipeline supports a shift toward a psychology that is less WEIRD and less ODD, but also more **WILD** (Newson et al., 2018). It enables the inclusion of participants ***W****orldwide*, in research conducted ***I****n situ* - that is outside the laboratory - informed by ***L****ocal* values and belief systems, and in ***D****iverse* samples that extend beyond university students or families living near research institutions. It is open source, based on video recordings that can be easily collected by anyone and pedagogically designed to support use by researchers with minimal programming experience. Thus, it has the potential to help make psychology a science grounded not only in the systematic observation of real-life behavior, but also one that is less elitist and marginalizing. No longer limited by WEIRD participants, theory, methods, or institutions (Oliveira & Baggs, 2023), we envision that pipelines of this kind, with their potential for universal availability and applicability, can support realization of diversity, equity, and inclusion in psychological research and theory.

Temporal Dynamics of Behavior

Classical observational studies quantified human activities and provided information about their frequencies and proportions; however, they often failed to capture the temporal dimension of unfolding behavior and social interactions. Averaging values over time, which reduces variability typically regarded as "noise," results in the loss of critical information about behavior and its interdependencies. By contrast, the "process-oriented" approach (van Geert & de Ruiter, 2022) does explicitly consider temporal properties, enabling observational measurement that captures temporal ordering and co-dependencies of unfolding behavior and dynamics of interaction through methods such as sequential (e.g., Bakeman & Gottman, 1997) or time series analysis (e.g., Guastello & Gregson, 2011). The growing availability of digital recordings has further facilitated the study of these dynamic, time-related aspects of behavior, making it possible to quantify change over time (Dishion & Granic, 2004); however, when such analyses rely on manual coding, they remain time-consuming and labor-intensive. CV-based measurement overcomes these limitations, allowing for precise and dense behavioral measurement beyond the scope of human coders by continuously estimating points of the human body in each frame of video recordings. Complemented with dynamical time series analyses, such as Recurrence Quantification Analysis (RQA, Webber & Zbilut, 1994) and Cross- RQA (CRQA, Marwan & Kurths, 2002), the possibility to discover temporal patterns of behaviors by investigating non-linear dynamics of multivariate continuous data is greatly expanded beyond the sequential categorical analyses (Dale et al., 2011).

Why Movement Matters: From Time Series to Understanding Behavior

CV-based motion tracking, as in traditional observational research, requires the specification of observation "targets", which remain inherently observer-dependent. Data captured on video are a source of thousands of variables that can be used to specify behaviors of interest. Without any constraints, facing these high-dimensional data puts us in a similar situation to a researcher without a theory nor a coding scheme observing a live situation. How should one decide what is interesting and worth studying without narrowing the scope of observation too much? The proposed pipeline supports three ways to focus observations: data-driven approaches that find recurring movement patterns automatically; theory-driven approaches that select measurements based on established research concepts (Maass et al., 2018); and experience-driven approaches that use researchers' and participants' insights to identify meaningful changes in how interactions unfold.

The theory-driven approach hinges on the dynamical systems view of behavior (Thelen & Smith, 1994), which assumes that every movement of a living organism is a manifestation of engagement in multiple 'projects' (i.e., goal-directed actions; Merleau-Ponty, 1945). Therefore the dynamics of the movement itself, and of the relations of this movement to other movements, is informative about these projects (Van Orden et al., 2003). Patterns of movement that recur are therefore informative about the stable organizations within the system itself or of the coupling of the system with the environment (Olthof et al., 2023). Take an undisclosed example of a moving human being: analyzing movements of the body parts, we would quickly note a curious, strong coupling between the movement of legs and arms. Such relative recurrent organization is an indication of the presence of behavior that is functional for the organism (in this case walking). Even without the first-hand knowledge of walking or the observation of walking in nature, the presence of such a pattern of movement, obtained via CV would indicate the importance of this behavior for a given organism.

At an interpersonal level, bodily movement coordination can – with some degree of certainty – reflect aspects of a relationship between people. For example: is the conversation they are having argumentative or affiliative? Is one person a leader or is it a symmetrical exchange (Paxton & Dale, 2017)? Relations among the selected aspects of the captured movement (such as synchrony, lag correlations, leader-follower dynamics) are informative about the interaction processes between two or more participants, instead of being focused on individual "behaviors" typically interpreted as actions characterizing a person. Capturing such dynamical structures underlying interpersonal interactions means we seek patterns of coordination sustained at multiple levels of organization (Kelso, 1995). The pipeline presented below exemplifies such particular dynamical systems approach to understanding behavior. However, it remains *theoretically open*, inviting other approaches, and ensures that researchers collect and curate privacy-protecting data tailored to their research questions, enabling the reuse of archived data in theory-driven, data-driven, and third-person experience-driven approaches.

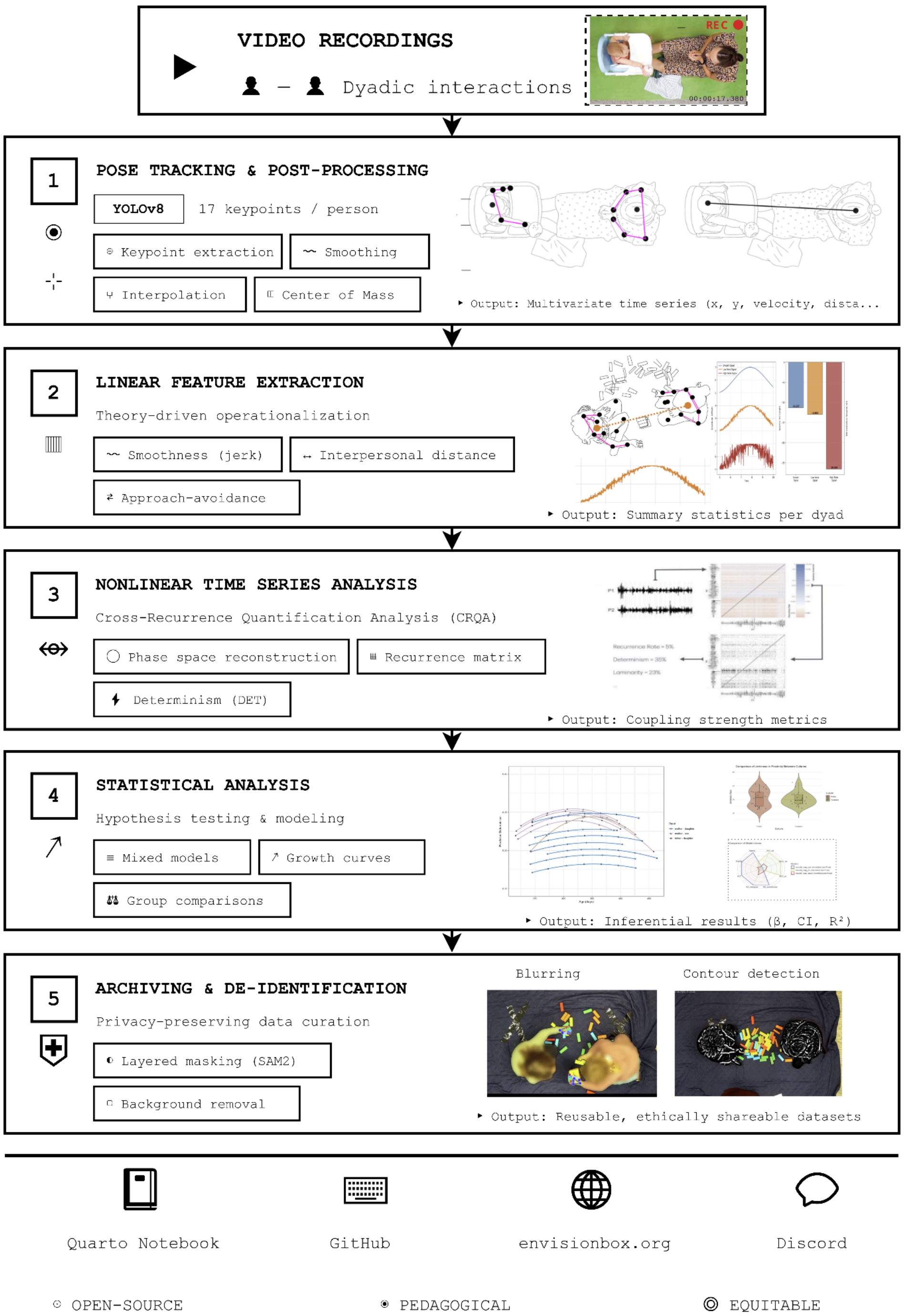
Workflow Overview

The workflow proceeds in five primary steps (shown in Figure 1). Step 1 covers

observation, applying CV to video data in order to acquire interaction-relevant time series (e.g., proximity at time *t*). The use of CV ensures a reproducible, quantitative approach to capturing observational data. Step 2 covers **(linear) feature extraction,** summarizing time series into single features representative of particular behavioral constructs. For our cases, we calculate the *smoothness* of an interaction by quantifying interpersonal distance and its abruptness of change over time (Case 1) and in Case 2 we assess coordination dynamics of interpersonal approach-avoidance behavior. Step 3 covers **nonlinear time-series analysis** for the study of coordination dynamics. Specifically, Cross-Recurrence Quantification Analysis may be adopted (CRQA; as demonstrated in our case examples) to assess bodily proximity changes between dyads, providing for each dyad features (e.g., determinism as a measure for dynamic predictability) that characterize the coordination dynamics. Once all metrics are calculated, we move to step 4: **statistical analysis.** This step covers statistical testing of the smoothness and non-linear analysis results. For the case studies, we assessed the changes in the smoothness of proximity measures across groups (Case 1) or age (Case 2). Finally, step 5 covers **archiving and deidentification strategies**. This step involves masking while preserving relevant kinematic information, ensuring that data are stored in an accessible, sustainable fashion that maximizes openness while ensuring privacy.

It is important to emphasize that at each step other choices concerning variables and underlying models can be made and/or other further analyses can be proposed. For example, after step 2 the resultant data can be used to assess overall turn-taking timing (Raffensperger et al., 2012) or movement synchrony (Hudson et al., 2023), instead of focusing on smoothness or subjecting it to non-linear time series analyses as in the cases described here The steps of the pipeline are therefore presented as generally as possible and the exact implementation can be found in the supplemental computational notebook.

Figure 1. *Overview of Workflow for Open Source Standardized Pipeline for Equitable Observations of Interactive Behavioral Dynamics*



Pipeline Details

To enhance reproducibility, we provide a Quarto Notebook that presents the methods and results in extended and fully reproducible format. This Notebook is an essential component of the present contribution as it allows researchers to implement (and adapt) an instantiation of our proposed pipeline in a step-by-step fashion. The Notebook includes details, elaborations and complete references[2]. The descriptions we provide here are abbreviated conceptual summaries of the phases of the procedure to be able to argue for the potential value of the current approach for an equitable behavior science.

Transparency and Openness

The current demonstrative position paper is not a confirmatory research contribution: The demonstrated analysis and the studies drawn upon are therefore not pre-registered. However, all code (processing, statistical analysis, masking) is fully reproducible and can be found on github, with the notebook as the key landing page providing a step-by-step technical demonstration with detailed conceptual explanations. Upon publication we will move our github materials to https://codeberg.org/, a non-profit GDPR-aligned transparent alternative. Materials used for this pipeline are explained in detail (e.g., computing infrastructure), and materials used in the studies are minimal (and described in the methods). Study protocols are described in the methods. All de-identified quantitative pose data and further derived data structures are provided on github. Masked recordings, provided with parental consent, will be deposited in the ***Anonymized*** Open Research Data Repository and may be shared with other researchers upon reasonable request once a data use agreement has been signed. During peer review the masked data snippets are available on a password-protected server (password: InterPerDynPipeline) so researchers get a preview of the dynamics that can be investigated with the corpus and accompanying dataset.

Step 1: Pose Tracking, Post-processing, Animating

[2] As a side note, page limits or reference limits lead to disparities between who is likely to be cited, and we think increasing extended reference sections to be future-proof practices for (society-run) journals (next to other practices in scholarly publishing such as increasing biblio-diversity and diamond open access publishing).

Step 1.1 Extracting Poses Per Frames (implementation e.g., YOLOv8). Pose tracking from video can be performed by any of the multiple tools available (such as, e.g., MediaPipe, OpenPose or YOLO). The choice for a particular pose model depends on many factors, such as camera angle availability, differences in research emphasis on whole-body versus hand movement, or the computational resources available. For the Cases described here, YOLOv8 (Jocher et al., 2023) was selected for its high consistency in pose tracking from a top-view camera angle. YOLO provides 17 pose keypoints per detected person which were filtered to ensure the quality of the tracking data. Specifically, we excluded duplicate detections of persons, as well as skeletons with excessive missing data. Particle detection of the skeleton often concerned mistrackings of individuals or tracking of the experimenter who is momentarily in the frame. The parameters that determine the exclusions are adjustable by the user. For specific information about the tracking procedure (e.g., list of keypoints), see the notebook.

Step 1.2. Post-processing to Time Series. Raw pose data were transformed to time series matrix format, which means for each frame we have a timestamp (in seconds), keypoint data, and derived measurements (e.g., speed wrist, center of mass). All keypoint data and derived measures were smoothed to reduce noise-related jitter in the measurements, and all derived measures such as speed were also smoothed (with a Savitzky-Golay filter, see notebook). Missing data points for particular frames were linearly interpolated to increase the quality of the tracking (YOLO may fail to track all key points across all frames). The derived variables were, for example, wrist speeds for left and right hand for both persons, or shoulder midpoint x,y locations. Furthermore, we derived variables such as the "center of mass" (COM) measure for each person, which is defined by the average horizontal and average vertical location of the upper body key points for each person (see Figure 1). Those variables allow for deriving the Euclidean distance between each person's center of mass coordinates.

Central to our main research questions involved extracting measures that track approach-avoidance behavior. Such behaviors could, in principle, be defined by an interpersonal distance measure, and in Case 1, we implemented this by measuring the Euclidean distance between person 1 and person 2's center of mass. The change of this distance measure over time was then administered to the smoothness analyses (Step 2). However, a simple distance measure cannot distinguish between *who* of the two persons is approaching or avoiding at a particular moment, losing information about coordination dynamics. Case 2 required deriving a measure that would capture this person-specific approach-avoidance. This “Social proximity” determines whether the person A approached/avoided B and/or B approached/avoided A at any particular time, while ignoring dynamic rearrangements (e.g., angle between the persons). We further differentiated the changes in the COM with respect to each time step (i.e. video frame), for each individual, which would later form input for their non-linear analysis of the interaction dynamics (Step 3). For a detailed explanation see the notebook.

Step 1.3 Video-Animations. The pipeline further provides animation code that is designed to allow for qualitative checking of the time series. The user can select variables that need to be plotted in the video that is then integrated with the tracked video. This way, the derived time series and measures can be compared directly to the behavior observed in the video, increasing qualitative understanding of the quantitative measures, further serving as a sanity check of preprocessing steps such as smoothing of time series (see here for details).

Step 2: Feature Extraction

The next step in the pipeline is to extract features from the time-series data. In principle, any features can be extracted and the choice is motivated by theoretical concerns and concrete research questions. For example, one can choose to determine features in a fully data-driven way (Christ et al., 2018), or one can focus on variables that summarize synchrony over time, or turn taking, or statistical measures of complexity. For the Case 1 analysis, we focus on operationalizing the concept of smoothness. We use *log dimensionless squared jerk* as a measure of smoothness (Hogan & Sternad, 2009). For Case 2, we generate features based on non-linear time series analyses which is described in the next step.

Step 3: Non-linear Time Series Analysis

In Case 2, the development of infant-caregiver interaction was studied using CRQA, which, as mentioned above, is a nonlinear time series analysis method for studying synchronization and the strength and direction of coupling dynamics between two systems (Marwan & Kurths, 2002; Shockley et al., 2002). Specifically, we used CRQA to quantify the coupling between infant and caregiver. To assess whether the overall strength of coupling changes over time, a generalized mixed model will be used, with the CRQA measure that quantifies coupling strength (determinism) as the dependent variable and age and dyad relation as predictors.

Using CRQA, the extent to which the dynamics of two different systems are coupled over time can be quantified by evaluating whether they occupy the same regions in a shared multidimensional phase space. In this space each coordinate represents a state, and a trajectory represents the evolution of system states over time. A phase space trajectory can be reconstructed from a single observed time series by creating delayed copies that represent so-called surrogate dimensions. Takens' theorem (Takens, 1981) guarantees the reconstructed trajectory will be topologically equivalent to the original, the important dynamical properties are recovered, not the exact same trajectory. The embedding procedure includes estimating two parameters, an embedding delay and the number of embedding dimensions. Using the same delay and embedding dimension for all time series enables constructing a cross-recurrence matrix for each dyad and comparisons among them (in Case 2 delay 838 and embedding dimension 3 were used, for details see notebook step 3).

Many different measures can be calculated from the cross-recurrence matrix, in the present study, the focus is on determinism (DET), defined as the proportion of the number of recurrent points that form diagonal lines of specified length to the total number of recurrent points. Diagonal lines indicate that a consecutive sequence of states that occurred in one time series is repeated in the other time series in almost exactly the same way. In other words, at some point during the observation period, both systems followed the same path through state space for a while. Therefore, determinism is generally considered a measure of the strength of the coupling between the two systems.

Step 4: Statistical Analysis

Step 4 consists of statistical testing of the observed phenomenon in order to assess a particular research question or test specific hypotheses regarding differences between individuals, sessions, or groups. It can also include model building such as mixed effects models or structural equations models. What we advise in this step is to follow best practices for conducting as well as reporting statistical analyses (Lakens, 2022). For our specific examples, in Case 1 and 2 we use a variety of common tests such as t-tests and linear mixed effects models (Case 1) and growth curve models (Case 2). While CV-based movement extraction allows for rich analyses of the temporal dynamics, in Case 1, we demonstrate an approach that aggregates measures across the interaction into a single summary statistic (yet still a temporal statistic; Abney et al., 2025). In Case 2, we demonstrate a more process-oriented approach that preserves even more of the temporal dynamics using non-linear time series analysis. Additional details of those methods are provided below.

Case 1 Analyses: Movement fluidity of sibling cooperation in multiple cultures. Cooperation, as 'doing things together', involves partners sharing a goal, and coordinating actions and intentions to achieve it. Its development is embedded in cultural contexts, and cultural differences in cooperation patterns should already be observable in childhood.

Children from Indigenous communities in the Americas have been observed to exhibit what has been called "fluid collaboration," functioning as a cohesive ensemble that integrates individual goals, and actions, anticipate each other's contributions, respond harmoniously to each other, contributing smoothly with the shared direction and rhythm of the group (Ruvalcaba & Rogoff, 2022). In contrast, Euro-American children tend to share tasks, leadership, and resist suggestions, reflecting a more *negotiation-based* form of interaction (Mejía-Arauz et al., 2018).

We will focus on studying cooperation in childhood through behavioral analysis of sibling interactions during a tower building task in two different cultures: Yurakaré Indigenous community of Bolivia and Polish urban culture. Yurakaré children grow up in mixed-age peer groups that foster active participation through cooperation and adaptive adjustment to others. Polish children represent a Western context characterized by child-centered, adult-organized activities that prioritize psychological autonomy. Our hypothesis is that Yurakaré children, as representatives of an Indigenous, cooperation-oriented culture, will demonstrate more fluid and smooth dynamics during the task than Polish children. Most research on early cooperation relies on pairings of same-age peers and overlooks the mixed-age groups that are typical of children's social environments in traditional societies. Our research addresses this gap.

The cross-cultural patterns of collaboration described above were studied using observational coding into mutually exclusive categories within specific time-segments. CV tools allow bottom-up, time-series analysis of interaction dynamics, enabling quantitative study of these phenomena. Thus, in the movement-based task of sibling pairs in the presented case study, we operationalize “fluid collaboration” in terms of the *smoothness of their movements*. But, given the high dimensional data extracted from the videos, we focused on two metrics one that is based on the smoothness of the individual approach/avoidance behavior (see Step 2 of the pipeline, explained in detail in our notebook) and one that is dyadic and based on the smoothness of the distance between the COM between the members of the dyad.

We observed 10 sibling pairs from each group (Yurakaré: *Mage* = 5.5 years, *SD* = 21.03 months, 70% female; Polish urban: *Mage* = 5.05 years, *SD* = 17.42 months, 45% female). Observations were made during a semi-structured play situation: in the natural environment of the Yurakaré community and laboratory conditions in Poland. Siblings were asked to build a tower together using 54 wooden blocks (see masked video in notebook). Each pair had 4 minutes to complete the task. This setup allowed for a systematic comparison of coordinated and mutually adjusted movements, using our pipeline to examine the smoothness of interaction while "doing things together". Interactions involving Polish siblings were recorded with an AXIS V5938 PTZ Network Camera, and Yurakaré interactions with a Nokia Lumia 1020 smartphone camera. The research was reviewed and approved by the Research Ethics ***Anonymized***, on May 12, 2021 for Yurakaré group and ***Anonymized*** Polish sample.

Note that for these analyses, we use a single summary statistic aggregating across the observed 4 minute period of interaction. For both of these metrics, lower values indicate more smooth movements and higher values indicate more jerkiness (i.e., less smooth) in the movements. For the individual level analysis of the smoothness of the individual approach/avoidance behavior, we fit a linear mixed model where the individual smoothness metric was the outcome variable, culture (Yurakaré or Polish) was the predictor variable, and we included a random intercept to account for individuals being nested within sibling pairs. Overall, the model suggests there was no evidence of an effect of culture on the smoothness of individual approach/avoidance movements, (β = 0.12, 95% *CI* [−0.53, 0.77], marginal R^2 = .004). For the Polish group, the estimated marginal mean was 38.3 (*SE* = 0.45, 95% *CI* [37.3, 39.2]). For the Yurakaré group, the estimated marginal mean was 39.5 (*SE* = 0.45, 95% *CI* [37.6, 49.5]). Figure 2 shows the overall pattern. In order to evaluate whether the results of these models were biased due to singular fit issues in estimating the random effects, we ran an equivalent linear regression including cluster robust standard error estimates which provided compatible results (see notebook for full details).

For the dyadic distance between center of mass analyses, we fit a simple linear regression with the smoothness of the distance between the dyad members' center of mass as the outcome variable and culture (Yurakaré vs Polish) as the predictor variable. Consistent with the prior analyses, we did not observe a significant effect for culture ($\beta = 0.15$, 95% *CI* [−0.81, 1.11], $R^2 = .006$). The estimated marginal mean for the Polish group was 39.0 (*SE* = 0.64, 95% *CI* [37.7, 40.3]), and for the Yurakaré group it was 39.3 (*SE* = 0.64, 95% *CI* [38.0, 40.6]).

Case 1: Discussion. We did not observe differences in the smoothness of movements between Yurakaré and Polish siblings. As was made transparent above, there are many possible ways one might have operationalized this psychological construct. Our choice might have not reflected other qualitative differences between populations. For example, the more extensive language use by Polish children might have constituted an additional constraint coordinating them and might have hidden the expected disfluencies. Our pipeline enables additional qualitative observations and quantitative coordination analyses based on other features in order to contribute to the discussion of the "smoothness" concept - e.g., its abandonment or redefinition. Thus, what appears as "smooth" adjustment might involve sudden behavioral shifts that the jerk measure may not adequately capture. These contexts also have a high degree of turn-taking such that when smoothness measures are aggregated across the entire period of observation, these measures may not quite reflect the complementarity of interaction that is turn-based (Fusaroli et al., 2014; Raffensperger et al., 2012). And lastly, we focused primarily on movements, but from a complex, dynamical systems perspective there are many relevant 'scales' of social interactions (Abney et al., 2025) to analyze and it could be that the "fluid collaboration" is more evident, for example, in the vocalizations or pliant leadership.

Case 2: Longitudinal Infant-Caregiver Interactions. Interactions between infants and their caregivers are essential for early developmental processes. Infants from an early age engage in coordinated interactions with caregivers and become sensitive if caregivers contingently respond to their behavior by 4 months (Beebe et al., 2016). Parents provide structure for such rhythmic interactions; they “frame” infants’ actions. Although timing matters in all social interactions, it is especially crucial for mother-infant interactions, where the focus is more on when things happen than what happens — the rhythm matters more than specific actions (Fogel, 1988). Caregiver-infant interactions involve multiple channels: vocalizations, gaze, and rhythmic body movements during back-and-forth sequences. These timing patterns reveal the structure of social connection without requiring identification of specific behaviors.

Timing patterns identified in research to date were identified predominantly on the basis of manual coding of behaviors cataloged in a coding scheme. CV-based measurement enables the study of the temporal organization of interpersonal interactions in a novel way. Its application will be illustrated using recordings from a longitudinal, semi-naturalistic observational study focused on development of pointing gesture in infant-caregiver interactions. Thirteen infant-caregiver pairs visited the child laboratory monthly for a total of seven visits. The infants' ages at the first visit ranged from 7 to 9 months (M_{age} = 236.4 days, SD = 17.26 days, 62% female). Parents were instructed to ‘play as usual,’ with a set of five age-appropriate toys while the infant was seated in a feeding chair facing the caregiver, who sat directly across the infant. The interaction was recorded using AXIS V5938 PTZ Network Camera. The research was reviewed and approved by the Research Ethics **Anonymized**, on February 11, 2022.

To analyze the development of parent-infant interactions, we fit a logistic (generalized) mixed model predicting the determinism (DET) measure obtained from CRQA for each dyad at each measurement occasion. The age of the infant (in days) was used as the time predictor and was centered on the average age of the sample at the first measurement occasion. The framework of generalized mixed models allows for the estimation of proportion data based on counts, where the denominator may vary from case to case (Renjaän et al., 2024). In the present dataset 2 dyads had 1 time series that was shorter than expected based on the selection criterion of 150 seconds, 4 dyads had one time series that required the removal of movement artefacts. These datapoints were set to missing values. Time series length will affect the number of recurrent points that can be detected; therefore, we used the number of recurrent points detected as a weight variable in the model (see supplementary data for details). A linear and quadratic effect of the time variable (age in days) were included in the model, as well as an interaction with a categorical variable indicating parent-infant relation; mother-daughter (*N* = 6), mother-son (*N* = 5), father-daughter (*N* = 1) with the latter as the reference category. The dyad ID was entered as a random effect in the model. The model predictions are displayed in Figure 2 The estimated marginal means at 236.4 days (the average at the first measurement occasion) for the father-daughter dyad (*M* = .39, 95% *CI* [.36, .42]) did not differ from the mother-daughter dyads (*M* = .36, 95% *CI* [.35, .37]) and mother-son dyads (*M* = .40, 95% *CI* [.39, .41]), but the mother-daughter and mother-son dyads did differ (*OR* = 0.86, z_{ratio} = -3.96, p_{Tukey} = .002). All interaction effects of the dyad type with linear and quadratic age effects were significant (for details see Table S1).

Case 2: Discussion. The observed pattern of growth in infant-caregiver coordination partly replicates findings from studies using manual coding; Tronick and Cohn (1989) found increasing coordination in infants 3, 6, and 9 months. Our sample included infants up to 13-14 months (in two cases up to 15 months), and after a similar initial increase, we observed a subsequent decline in strength of the coupling. Given the small sample, these findings should be interpreted cautiously. Notably, peak infant-caregiver coordination occurred around 10-11 months, a period of major social-cognitive developments like joint attention (Tomasello, 2018).

We also observed gender differences consistent with findings from previous studies using manual coding. Specifically, prior research has shown that at six and nine months, mother-son pairs spend more time in coordinated states than mother-daughter pairs (Tronick & Cohn, 1989), and that at six months mother-son dyads exhibit higher synchrony scores than mother-daughter dyads (Weinberg et al., 1999). Again, given our small sample size, the observed gender differences in the level of determinism should be treated cautiously. Tentatively, mothers and sons track each other's behavior more precisely and sustain structured interactions longer.

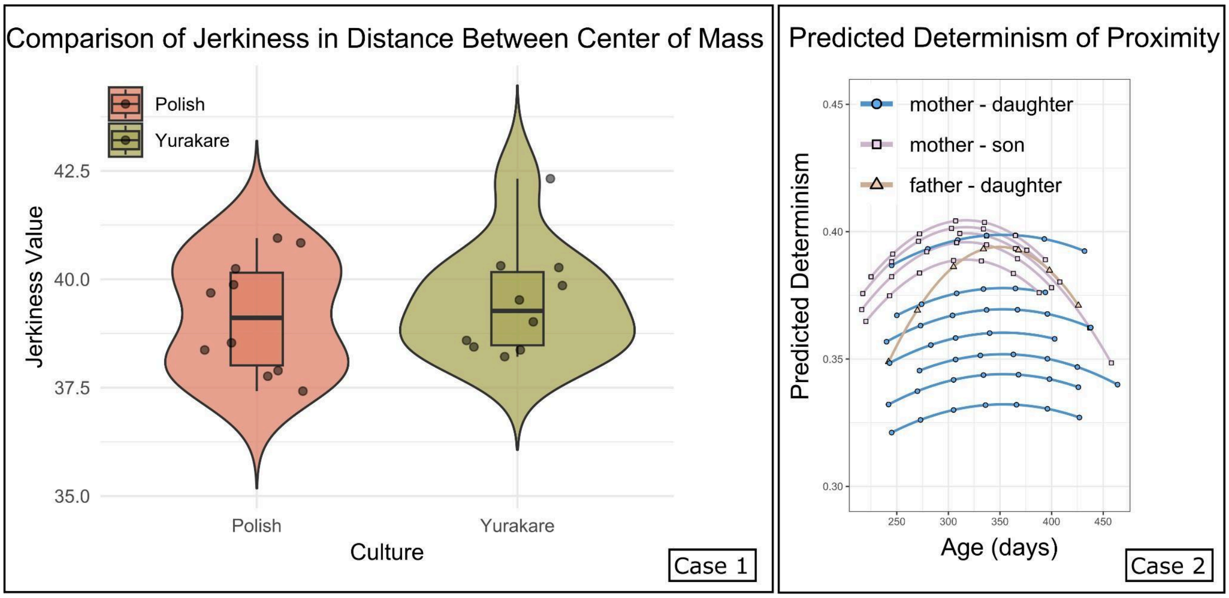


Figure 2. *Combined results for the case studies: Smoothness (Case 1) and CRQA (Case 2)*

Note. Case 1 (left plot): Cross-cultural Comparisons of Individual Jerkiness of Distance between Center of Mass (B). Case 2 (right plot): Predicted changes in Determinism by Age and type of Dyad. Predicted values are based on a logistic mixed model with linear and quadratic time predictors interacting with dyad type.

Step 5 - Protocols for Archiving Sensitive Data Using Masking Strategies

Behavioral video data offers significant scientific value for understanding human interaction but presents serious ethical challenges, particularly with machine learning adoption. Movement data retains ethically significant information about culture, affect, health, age, and neurodevelopment even after facial de-identification, reflecting meaningful social patterns related to identity and relationships that require protection rather than erasure (Light et al., 2024). Video observation captures embodied and temporal dynamics, yet this richness renders data highly identifying, particularly for vulnerable populations like children or culturally specific communities, while conventional approaches often restrict access entirely, creating data silos that hinder reproducibility and collaboration (Landi et al., 2020).

Concerning layered archiving of time series and masked videos, the pipeline produces two complementary outputs for archiving and sharing, each serving distinct purposes in the data reuse ecosystem. Time series data as the primary shareable output. At the foundational level, computer vision tools like YOLO extract movement data as high-frequency time series of body landmark positions and trajectories (Gao et al., 2025). These time series—including keypoint coordinates, derived measures (e.g., velocity, interpersonal distance, center of mass), and summary statistics (e.g., smoothness, CRQA metrics)—constitute fully de-identified behavioral data. Because no visual information is retained, these data carry minimal re-identification risk while preserving the full analytic value of the original recordings. Researchers seeking to replicate or extend analyses should use these time series directly; they contain all information necessary to reproduce the statistical results reported in a study.

When visual context is necessary to interpret the behavior—for example, to understand the physical setting, verify tracking quality, or conduct supplementary manual coding—a second archiving level preserves movement while obscuring identifiable features. This is achieved through targeted blurring, pixelation, conversion to silhouettes or skeletal figures, and background removal using segmentation models like SAM2 (Segment Anything Model; Ravi et al., 2024) and open-source tools such as MaskAnyone CLI (Owoyele et al., 2024). These transformations offer a compromise that maintains interactional content while achieving ethically acceptable de-identification levels.

Importantly, masked videos are not intended as input for re-running pose estimation. Preliminary evaluation indicates that applying pose estimation models to masked videos yields detectably different outputs compared to original recordings—a finding we interpret as evidence that the de-identification transformations effectively disrupt the visual features these models rely on. This degradation in model performance does not compromise data reusability, because the high-fidelity time series extracted from the original videos remain the primary resource for quantitative reanalysis. The masked videos and the time series data thus serve complementary roles: the former supports qualitative understanding and transparency, while the latter enables rigorous, reproducible secondary analysis.

Regarding selecting de-identification levels, the layered protocol enables researchers to select de-identification levels based on ethical review, participant consent, and intended use. For maximum openness, time series data can often be shared with minimal restrictions. For masked videos, the appropriate level of transformation (e.g., blur vs. silhouette vs. skeleton) depends on the population studied, the sensitivity of the behavioral context, and the consent obtained. The recommendation is that researchers consult with ethics boards and, where possible, with participant communities to determine acceptable levels of visual detail.

In terms of equity and the future of behavioral data sharing, this archiving approach enables broader, fairer participation by lowering re-identification risks while preserving analytic properties. It facilitates data reuse, supports replication, and reduces inequities in who gets studied. Ethical de-identification is not a constraint on behavioral science—it is a condition for its equitable future. It is important to reflect on the challenges encountered in fully archiving the raw videos,as the original study did not have signed agreements from all participants and because of the sensitive nature of the videos containing minors, masking serves as an intermediate step in the archiving workflow. Support from institutional repositories and pre agreed consent forms are practical strategies for granting fuller access. For now the CSV files and masked video snippets have been curated carefully to reflect the intended spirit of the archiving workflow-to be as open as possible, as closed as necessary.

The future of the pipeline: A community approach

The current technical demonstration of the pipeline is a starting point for equitable innovations in methodology. The pipeline should evolve from here, in part by project and aims of the current team, but also by recruiting community involvement. Some of the features we think can be further profitably integrated are video tutorials, specification of learning routes with open access learning materials related to components of the pipeline (e.g., R, Python, Kinematics, Time series, Dynamical Systems). Furthermore the notebook can be improved itself, with interactive FAQ, but also by more interactive graphical-user-interface (GUI) elements that invite researchers to connect their aggregate findings to the specific situations in the interactions (Miao et al., 2025). To increase community involvement in our notebook and future versions, the notebook will be staged (after review) as a module on non-profit open science platform www.envisionbox.org.

There we also reserved two channels on their forum so that users and developers can discuss specifics of using and improving the pipeline (https://discord.gg/GQmwYmS4). In this way, future versions or new derivative pipelines with more contributors will more likely come about through community effort (for similar approaches see Matthis, 2025).

Discussion

Our pipeline addresses traditional observational research limitations by making behavioral analysis accessible and equitable. The widespread availability of recording tools enables researchers to collect naturalistic data without high-tech laboratories, while the proposed open-source, pedagogically oriented design accommodates users with minimal programming expertise. Combined with masking procedures for privacy protection, this supports diverse populations and researchers previously excluded from such studies. What once required months of manual coding can now be processed in hours, making large-scale observational research feasible and overcoming traditional barriers of cost, complexity, and time. Rather than using predetermined coding schemes, our theory-driven approach examines emergent movement patterns. We operationalized constructs like movement fluidity through smoothness measures and coordination through coupled changes in center of mass. This creates opportunities to discover behavioral patterns that might not align with established frameworks.

Movement tracking reveals behavioral dynamics across multiple timescales, from micro-level coordination to developmental changes over months. This process-focused analysis aligns with dynamical systems perspectives and enables previously impractical questions about how behavioral patterns vary across populations, developmental stages, and cultural contexts which is essential for developing much-needed psychological theories that target humanity in its diversity.

Constraints on Generality and Other Limitations

Due to small sample size, the case studies presented in our paper serve primarily an illustrative purpose, rather than providing conclusive results, and thus have inherent limitations in terms of generalizability. At the same time, given the characteristics of the presented pipeline, we consider its usability to be relatively universal, and our contribution to confirmatory research lies in the pipeline's potential of future reuse. Potential users do need to obtain participants' consent to be recorded, and consent for their data to be masked and archived for a certain amount of time.

CV-augmented behavioral analysis raises complex ethical questions that extend beyond traditional research ethics. While seemingly neutral, CV methods and workflows can encode sensitive (mis)information about health, emotional state, and cultural background (Buolamwini & Gebru, 2018) and developments in computer vision are often used for questionable surveillance practices (Kalluri et al., 2025). While potentially reducing labor associated with observations, models still need to be trained on often human-annotated data. Furthermore, masking protocols provide multiple levels of de-identification, but the fundamental challenge remains: how do we balance scientific utility with participant privacy while taking into account that sophisticated algorithms of tomorrow might retrieve information that was masked by technology we have today?

We must also resist the temptation to view automated analysis as inherently objective. CV systems embody the biases of their training data and the assumptions of their developers; they are as accurate and equitable as data on which they were trained (Fabbrizzi et al., 2022). Awareness of such biases has prompted efforts to address them, including the development of mitigation strategies such as the collection of bias-controlled datasets. It is essential that the application of such methods continues to be grounded in cross-cultural research and collaboration, along with a critical validation of the perspectives we adopt. More generally, a key feature of observational coding schemes used in manual analyses is their reliance on the "unit of meaning" rule (Lamb, 1979), according to which the coding system identifies behavior units typically used by the focal interactants. Although this approach entails a degree of researcher subjectivity and potential bias, CV–based tracking lacks such interpretive grounding. Consequently, validating CV-based results through comparisons with participant and researcher ratings remains an important and urgent task, as has been done in validation studies comparing CV-based measures with motion captures (Needham et al., 2021).

Video processing requires significant computational resources, making environmental considerations important. As these methods become widespread, researchers should minimize environmental impact while maximizing scientific benefit. Reusing masked data helps offset these costs by reducing redundant processing.

Another caveat should be mentioned concerning the current pipeline: there is an inherent limit to the degree that "feature" extraction can capture the significance of social interactions. Collapsing the temporal dimension into single-point estimates makes it easier to use traditional statistical tools, but it then simplifies the context-dependent and multi-scale unfolding of interactional dynamics that participants engage in. The researcher is ultimately tasked with balancing the inherently meaningful qualitative observations of (reported) lived experiences, against the quantitative measurements that aim to capture something of significance. One promising way forward is to design analyses that retain temporal information rather than collapsing it entirely, for instance, through windowed analyses and dynamical systems approaches that directly characterize recurring structures. These approaches make it possible to identify recurrent patterns or changes over multiple timescales while still grounding them in the specific observable phenomena (Gorman et al., 2020; van Eijndhoven et al., 2023). In this way, integrating qualitative and quantitative functions of measurement such that researchers are increasingly able to perceive new significances in social interactions through technology and formally specify the recurrent aspects of significant moments through technology.

Several promising developments could extend our approach. Real-time analysis capabilities would enable new forms of intervention research where participants receive immediate feedback about their interaction patterns (Wiltshire et al., 2024). Integration with other data streams such as physiological measures, environmental sensors, linguistic analysis could provide richer insights into the constraints and functions of behavioral dynamics (Alviar et al., 2023). We envision combining automated analysis with systematic inclusion of participant perspectives. Further, precision movement data provided by CV, when analyzed within Empirical Dynamic Modeling (Munch et al., 2023), opens a new avenue for investigating relationships between, for example, caregiver behavioral predictability and child development. The relationship between movement dynamics and established psychological constructs remains an active area of investigation (Warren, 2006). Future research should examine how patterns detected through automated analysis relate to traditional observational measures while remaining open to the possibility that movement-based approaches might reveal novel aspects of human behavior.

Beyond the traditional research sphere we are making this pipeline accessible to early career researchers and citizen scientists via bootcamps and summer schools.

Conclusions

Our open-source pipeline is designed to facilitate the collection, analysis, dissemination, and reuse of observational data in an equitable manner. Its use can not only help ground psychological research more firmly in observation of behavior, but also significantly broaden the diversity of studied populations. The pipeline advances traditional application of observational methods and creates opportunities for both researchers and participants from historically marginalized groups to contribute to the development of psychology. While we illustrate its usability within process-oriented and dynamic systems framework, the pipeline is adaptable to a range of theoretical approaches.

References

Abney, D. H., Fausey, C. M., Suarez-Rivera, C., & Tamis-LeMonda, C. S. (2025). Advancing a temporal science of behavior. *Trends in Cognitive Sciences*, *29*(12), 1109–1119. https://doi.org/10.1016/j.tics.2025.05.010

Alviar, C., Kello, C. T., & Dale, R. (2023). Multimodal coordination and pragmatic modes in conversation. *Language Sciences*, *97*, 101524. https://doi.org/10.1016/j.langsci.2022.101524

Bahlai, C., Bartlett, L., Burgio, K., Fournier, A., Keiser, C., Poisot, T., & Whitney, K. (2019). Open science isn’t always open to all scientists. *American Scientist*, *107*(2), 78. https://doi.org/10.1511/2019.107.2.78

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge University Press.

Barbosa, A. V., & Vatikiotis-Bateson, E. (2006). Video tracking of 2D face motion during speech. *2006 IEEE International Symposium on Signal Processing and Information Technology*, 791–796. https://doi.org/10.1109/ISSPIT.2006.270906

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A systems view of mother-infant face-to-face communication. *Developmental Psychology*, *52*(4), 556–571. https://doi.org/10.1037/a0040085

Berk, S. L., & Verghese, A. (1990). General Appearance. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Butterworths. http://www.ncbi.nlm.nih.gov/books/NBK330 /

Bernard, H. R. (2017). *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Bloomsbury Publishing PLC.

Brauner, E. (2018). Coding interaction: A technical introduction. In *The Cambridge handbook of group interaction analysis* (pp. 165–190). Cambridge University Press. https://doi.org/10.1017/9781316286302.010

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Chai, J., Zeng, H., Li, A., & Ngai, E. W. T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, *6*, 100134. https://doi.org/10.1016/j.mlwa.2021.100134

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh – A Python package). *Neurocomputing*, *307*, 72–77. https://doi.org/10.1016/j.neucom.2018.03.067

Dale, R., Warlaumont, A. S., & Johnson, K. L. (2023). The fundamental importance of method to theory. *Nature Reviews Psychology*, *2*(1), 55–66. https://doi.org/10.1038/s44159-022-00120-5

Dale, R., Warlaumont, A. S., & Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos*, *21*(04), 1153–1161. https://doi.org/10.1142/S0218127411028970

Darwin, C. (1872). *The expression of the emotions in man and animals* (pp. vi, 374). John Murray. https://doi.org/10.1037/10001-000

De Barbosa, V. (2017). *FlowAnalyzer*. https://www.cefala.org/FlowAnalyzer/

De Jaegher, H. (2021). Loving and knowing: Reflections for an engaged epistemology.

Phenomenology and the Cognitive Sciences, *20*(5), 847–870. https://doi.org/10.1007/s11097-019-09634-5

Dishion, T. J., & Granic, I. (2004). Naturalistic observation of relationship processes. In *Comprehensive handbook of psychological assessment, Vol. 3: Behavioral assessment* (pp. 143–161). John Wiley & Sons, Inc.

Doliński, D. (2018). Is Psychology still a science of behaviour? *Social Psychological Bulletin*, *13*(2), Article 2. https://doi.org/10.5964/spb.v13i2.25025

Fabbrizzi, S., Papadopoulos, S., Ntoutsi, E., & Kompatsiaris, I. (2022). *A survey on bias in visual datasets* (arXiv:2107.07919). arXiv. https://doi.org/10.48550/arXiv.2107.07919

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157. https://doi.org/10.1016/j.newideapsych.2013.03.005

Gao, Z., Chen, J., Liu, Y., Jin, Y., & Tian, D. (2025). A systematic survey on human pose estimation: Upstream and downstream tasks, approaches, lightweight models, and prospects. *Artificial Intelligence Review*, *58*(3), 68. https://doi.org/10.1007/s10462-024-11060-2

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring real-time team cognition during team training. *Human Factors*, *62*(5), 825–860. https://doi.org/10.1177/0018720819852791

Guastello, S. J., & Gregson, R. A. M. (Eds.). (2011). *Nonlinear dynamical systems analysis for the behavioral sciences using real data*. CRC Press.

Hogan, N., & Sternad, D. (2009). Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *Journal of Motor Behavior*, *41*(6), 529–534. https://doi.org/10.3200/35-09-004-RC

Hudson, D., Wiltshire, T. J., & Atzmueller, M. (2023). multiSyncPy: A Python package for assessing multivariate coordination dynamics. *Behavior Research Methods*, *55*, 932–962. https://doi.org/10.3758/s13428-022-01855-y

Jersild, A. T., & Meigs, M. F. (1939). Direct observation as a research method. *Review of Educational Research*, *9*(5), 472–482. https://doi.org/10.2307/1167746

Jocher, G., Chaurasia, A., & Qiu, J. (2023). *YOLO by Ultralytics* (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

Kalluri, P. R., Agnew, W., Cheng, M., Owens, K., Soldaini, L., & Birhane, A. (2025). Computer-vision research powers surveillance technology. *Nature*, *643*(8070), 73–79. https://doi.org/10.1038/s41586-025-08972-6

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. The MIT Press.

Lakens, D. (2022). *Improving Your Statistical Inferences*. https://doi.org/10.5281/ZENODO.6409077

Lamb, M. E. (1979). Issues in the study of social interaction: An introduction. In M. E. Lamb, S. J. Suomi, & G. R. Stephensen (Eds.), *Social interaction analysis: Methodological issues* (pp. 1–10). University of Wisconsin Press.

Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., da Silva Santos, L. O. B., & Roos, M. (2020). The "A" of FAIR – As open as possible, as closed as necessary. *Data Intelligence*, *2*(1–2), 47–55. https://doi.org/10.1162/dint_a_00027

Latif, N., Barbosa, A. V., Vatiokiotis-Bateson, E., Castelhano, M. S., & Munhall, K. G. (2014). Movement coordination during conversation. *PLOS ONE*, *9*(8), e105036. https://doi.org/10.1371/journal.pone.0105036

Light, L. L., Panicker, S., Abrams, L., & Huh-Yoo, J. (2024). Ethical challenges in the use of digital technologies in psychological science: Introduction to the special issue. *American Psychologist*, *79*(1), 1–8. https://doi.org/10.1037/amp0001286

Maass, W., Parsons, J., Purao, S., Storey, V., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems*, *19*(12). https://doi.org/10.17705/1jais.00526

Marwan, N., & Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A*, *302*(5–6), 299–307. https://doi.org/10.1016/S0375-9601(02)01170-2

Matthis, J. (2025). *The FreeMoCap Project*. https://freemocap.org/

Mejía-Arauz, R., Rogoff, B., Dayton, A., & Henne-Ochoa, R. (2018). Collaboration or negotiation: Two ways of interacting suggest how shared thinking develops. *Current Opinion in*

Psychology, *23*, 117–123. https://doi.org/10.1016/j.copsyc.2018.02.017

Merleau-Ponty, M. (1945). *Phenomenology of Perception*. Routledge.

Miao, G. Q., Trujillo, J., Bulls, L. S., Thornton, M. A., Dale, R., & Pouw, W. (2025). *DIMS* dashboard for exploring dynamic interactions and multimodal signals. In A. Ruggeri, D. Barner, C. Walker, & N. Bramley (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Munch, S. B., Rogers, T. L., & Sugihara, G. (2023). Recent developments in empirical dynamic modelling. *Methods in Ecology and Evolution*, *14*(3), 732–745. https://doi.org/10.1111/2041-210X.13983

Needham, L., Evans, M., Cosker, D. P., Wade, L., McGuigan, P. M., Bilzon, J. L., & Colyer, S. L. (2021). The accuracy of several pose estimation methods for 3D joint centre localisation. *Scientific Reports*, *11*(1), 20673. https://doi.org/10.1038/s41598-021-00212-x

Newson, M., Buhrmester, M., Xygalatas, D., & Whitehouse, H. (2018). Go WILD, Not WEIRD. *Journal for the Cognitive Science of Religion*, *6*(1–2), 80–106. https://doi.org/10.1558/jcsr.38413

Oliveira, G. S. de, & Baggs, E. (2023). Psychology's WEIRD problems. *Elements in Psychology and Culture*. https://doi.org/10.1017/9781009303538

Olthof, M., Hasselman, F., Oude Maatman, F., Bosman, A. M. T., & Lichtwarck-Aschoff, A. (2023). Complexity theory of psychopathology. *Journal of Psychopathology and Clinical Science*, *132*(3), 314–323. https://doi.org/10.1037/abn0000740

Owoyele, B. A., Schilling, M., Sawahn, R., Kaemer, N., Zherebenkov, P., Verma, B., Pouw, W., & Melo, G. de. (2024). *MaskAnyone toolkit: Offering strategies for minimizing privacy risks and maximizing utility in audio-visual data archiving* (arXiv:2408.03185). arXiv. https://doi.org/10.48550/arXiv.2408.03185

Paxton, A., & Dale, R. (2017). Interpersonal movement synchrony responds to high- and low-level conversational constraints. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01135

Raffensperger, P. A., Webb, R. Y., Bones, P. J., & McInnes, A. I. (2012). A simple metric for turn-taking in emergent communication. *Adaptive Behavior*, *20*(2), 104–116.

https://doi.org/10.1177/1059712311421831

Rai, T. S., & Fiske, A. (2010). ODD (observation- and description-deprived) psychological research. *Behavioral and Brain Sciences*, *33*(2–3), 106–107. https://doi.org/10.1017/S0140525X10000221

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., & Feichtenhofer, C. (2024). *SAM 2: Segment anything in images and videos* (arXiv:2408.00714). arXiv. https://doi.org/10.48550/arXiv.2408.00714

Renjaän, D., Vervoort, L., Ha, T., Hasselman, F., & Otten, R. (2024). Emotion coupling across socialization contexts in adolescence: Differences in parent–child and peer interactions. *Developmental Psychology*. https://doi.org/10.1037/dev0001865

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, *5*(1), 2–14. https://doi.org/10.1207/S15327957PSPR0501_1

Ruvalcaba, O., & Rogoff, B. (2022). Children's fluid collaboration versus managing individual agendas: Cultural differences in pair programming. *Journal of Applied Developmental Psychology*, *81*, 101438. https://doi.org/10.1016/j.appdev.2022.101438

Shockley, K., Butwill, M., Zbilut, J. P., & Webber Jr, C. L. (2002). Cross recurrence quantification of coupled oscillators. *Physics Letters A*, *305*(1–2), 59–69. https://doi.org/10.1016/S0375-9601(02)01411-1

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical systems and turbulence, warwick 1980* (Vol. 898, pp. 366–381). Springer. http://dx.doi.org/10.1007/BFb0091924

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. The MIT Press.

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift Für Tierpsychologie*, *20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences*, *115*(34), 8491–8498. https://doi.org/10.1073/pnas.1804761115

Tronick, E. Z., & Cohn, J. F. (1989). Infant-mother face-to-face interaction: Age and gender differences in coordination and the occurrence of miscoordination. *Child Development*, *60*(1), 85–92. https://doi.org/10.2307/1131074

van Eijndhoven, K., Wiltshire, T. J., Hałgas, E. A., & Gevers, J. M. P. (2023). A methodological framework to study change in team cognition under the dynamical hypothesis. *Topics in Cognitive Science*, *n/a*(n/a), 1–19. https://doi.org/10.1111/tops.12685

van Geert, P., & de Ruiter, N. (2022). *Toward a process approach in Psychology: Stepping into Heraclitus' River*. Cambridge University Press; Cambridge Core.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*(3), 331–350. https://doi.org/10.1037/0096-3445.132.3.331

Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, *113*(2), 358–389. https://doi.org/10.1037/0033-295X.113.2.358

Webber, C. L., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *76*(2), 965–973. https://doi.org/10.1152/jappl.1994.76.2.965

Weinberg, M. K., Tronick, E. Z., Cohn, J. F., & Olson, K. L. (1999). Gender differences in emotional expressivity and self-regulation during early infancy. *Developmental Psychology*, *35*(1), 175–188. https://doi.org/10.1037/0012-1649.35.1.175

Wiltshire, T. J., van Eijndhoven, K. H. J., Halgas, E., & Gevers, J. M. P. (2024). Prospects for augmenting team interactions with real-time coordination-based measures in human-autonomy teams. *Topics in Cognitive Science*, *16*(3), 391–429. https://doi.org/10.1111/tops.12606