A systematic investigation of gesture kinematics in evolving manual languages in the lab

Wim Pouw1, Mark Dingemanse1,3, Yasamin Motamedi4, & Asli Ozyurek1,2,3

1 Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen

2 Institute for Psycholinguistics, Max Planck Nijmegen

3 Center for Language Studies, Radboud University Nijmegen

4 Centre for Language Evolution, University of Edinburgh

Author note

This work is supported by a Donders Fellowship awarded to Wim Pouw and Asli Ozyurek and is supported by the Language in Interaction consortium project ‘Communicative Alignment in Brain & Behavior’ (CABB).

Correspondence concerning this article should be addressed to Wim Pouw, Montessorilaan 3, 6525 HR Nijmegen. E-mail: [w.pouw@psych.ru.nl](mailto:w.pouw@psych.ru.nl)

Abstract

Silent gestures consist of complex multi-articulatory movements, but are now primarily studied through categorical coding of the referential gesture content. The relation of linguistic content with continuous kinematics is therefore poorly understood. Here we reanalyzed the video data from a gestural evolution experiment (Motamedi et al. 2019), which showed increases in the systematicity of gesture content over time. We applied computer vision techniques to quantify and analyze the kinematics of the original data, demonstrating that gestures become more efficient and less complex in their kinematics over generations of learners. We further detect systematicity of gesture form on the level of their interrelations, which directly scales with the systematicity obtained on semantic coding of the gestures. Thus from continuous kinematics alone we can tap into linguistic aspects that were previously only approachable through categorical coding of meaning. Finally, going beyond issues of systematicity we show how unique gesture kinematic dialects emerged over generations, as isolated chains of participants gradually diverged over iterations from other chains. We thereby show how gestures can come to embody the linguistic system at the level of interrelationships between communicative tokens, which should calibrate our theories about form and linguistic content.

*Keywords:* language evolution, silent gesture, kinematics, systematicity

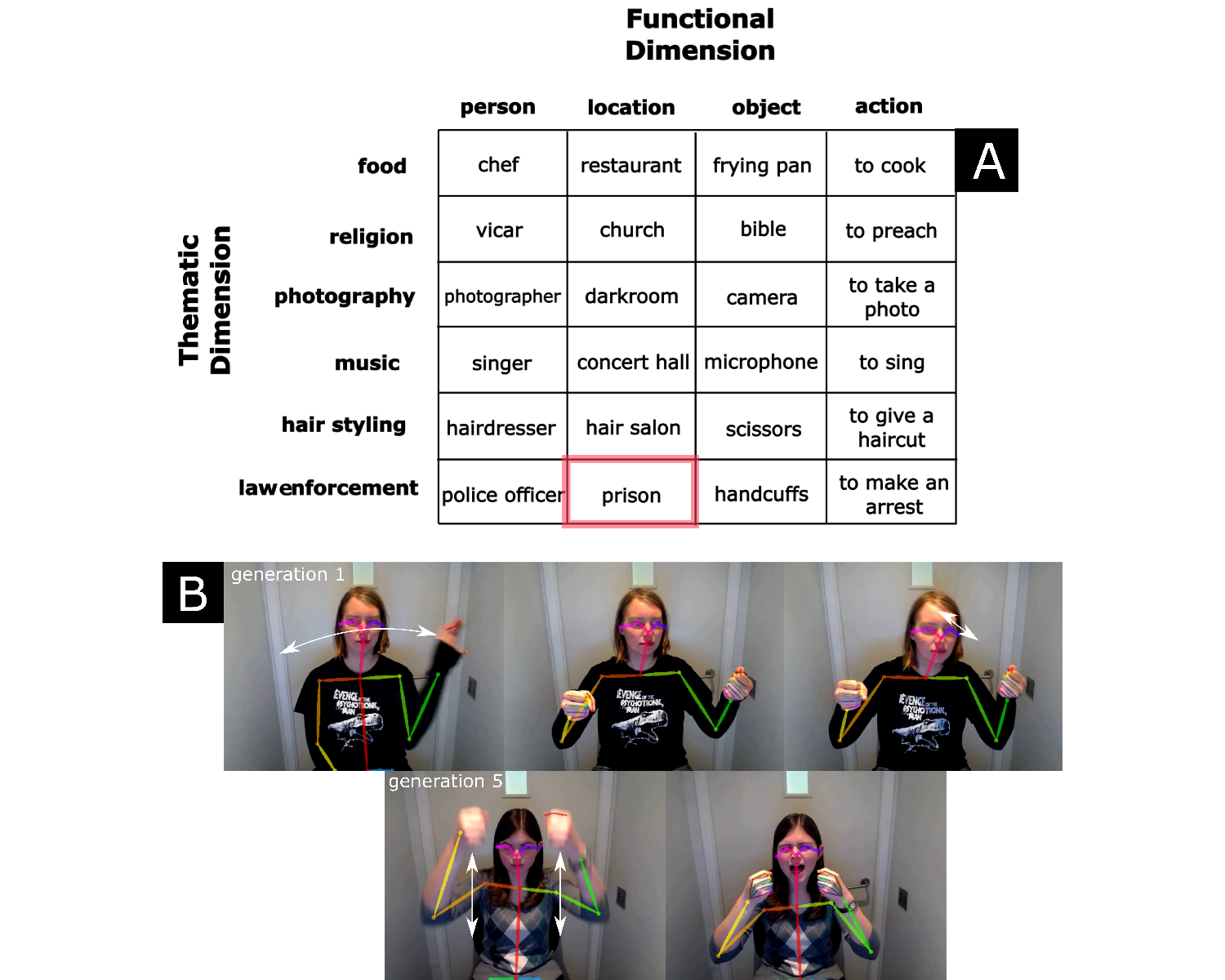
*Word count:* X

A systematic investigation of gesture kinematics in evolving manual languages in the lab

All known natural languages combine discrete categorical elements with continuous and dynamic properties (Bolinger, 1968). For a long time, the study of human communicative behavior has focused on aspects that best yield to analysis in terms of discrete categories such as lexical items, phonological building blocks, semantic categories, and their combinatorial properties. At the same time, language use is widely acknowledged to also feature more gradient and continuous streams of behavior that do not always easily yield to an analysis in terms of discrete symbol systems (Enfield, 2009; Kendon, 2004). Here we investigate whether kinematic measures directly derived from continuous manual movements can capture the meaning space they are designed to communicate. Using communicative silent gestures as a test case, we show how continuous movements can be studied as evolving dynamic systems.  
Manual gestures are seldomly *semiotically* studied based on the measurable part of a signal’s form, namely manual and whole-body postures in movement (i.e., the kinematic level). Instead gestures are mostly studied as already categorizable expressions by researchers inferring meaning from their form (McNeill, 2005). As such the kinematics are in one sense reduced (from continuous to discrete tokens) and in another sense enriched (from movement to message) with meanings that are projected onto them by human arbitrators. Here we study more abstract aspects of gestural systems through kinematic analysis. Our aim is to show that we do not need to leave the domain of form to observe the emergence of systematic properties. We suggest that a signal’s form, when studied in relation to other forms, can provide information about its linguistic properties that can complement or supplement cues that take into account conventional denotation and contextual information. Such systematicity in form —where low-level properties can serve as cues to higher-level regularities— is known from work on lexical classes in signed and spoken languages (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Padden et al., 2013). To observe it in existing linguistic systems, we can rely on conventional meanings and existing syntactic and semantic categories. Here we aim to capture its emergence in rich kinematic signals as they evolve over time. In doing so, we want to contribute to an understanding of how communicative signs come to function as interrelated parts of complex dynamic systems: relatively stable ways of signaling that form higher order structural wholes (Dale & Kello, 2018; Rączaszek-Leonardi & Kelso, 2008).  
The sense-making process of individual forms becoming parts of structural wholes is essentially simulated in iterated learning experiments (Kirby, Griffiths, & Smith, 2014; Motamedi, Schouwstra, Smith, Culbertson, & Kirby, 2019). In such experiments, agents are tasked to learn a novel set of signals. These signals are iteratively transmitted to later generations and/or used in communication by later generations (iterated learning + communication), where, over many cycles of learning and use, they are affected by various transmission biases (e.g., Christiansen & Chater, 2016; Enfield, 2014). Processes of iterated learning and communication can simulate how structural properties such as systematicity, learnability, and compositionality evolve from simpler communication systems. In such simulations, communicative tokens undergo cultural evolution constrained by population dynamic properties such as historicity (the system is constrained by past contingencies) and adaptivity (the system adapts in service of its informative goals). Such population dynamics must have played out over long temporal and vast population scales, but through these iterated learning paradigms such processes are to some degree brought under experimental control. These evolving or emerging communicative systems can be constituted by a variety of different signal media, from simple discrete symbol sequences to more challenging continuous acoustic signals (Cornish, Dale, Kirby, & Christiansen, 2017; Ravignani, Delgado, & Kirby, 2016; Verhoef, Kirby, & de Boer, 2016).

## 0.1 Current case study

Most iterated learning studies have simulated the cultural evolution of sign systems using easily discretized word-like written forms (Verhoef, Kirby, & de Boer, 2014; Verhoef et al., 2016). While this has made it easy to operationalize measures like compressibility, systematicity, and expressivity, it has done so at the expense of ecological validity. After all, the embodied semiotic resources that all known natural languages rely on are not carved into a predefined discrete symbol system like the Latin alphabet; instead, users and analysts of language alike must derive everything they know about linguistic systems from biological signals that are fundamentally dynamic and continuous (Pattee & Rączaszek-Leonardi, 2012).  
Iterated learning work has only recently moved into the area of continuous signals. The first adaptations focused on the emergence of phonological organization in continuous acoustic signals (Verhoef et al. 2014, 2016). Such signals do not afford a lot of semantic expressivity, so recent work has extended to the more daunting area of continuous multi-articulator signals in the form of manual depictions, or silent gestures (Motamedi et al., 2019). This experimental work studied the transmission of silent gestures created to communicate 24 concepts along two broad semantic dimensions: theme (e.g., food, religion) and function (e.g., person, location) (see Figure 1).

Figure 1. Concepts to be conveyed in gesture in Motamedi et al. 2019  
 *Note Figure 1*. A) The concepts and categories that were used in the original experiment are shown. B) Example of silent gestures depicting the concept “prison”. In the first generation, a drawn out multicomponent silent gesture is shown (multiple arm movements and head movement), while in the final generation 5 a much more simpler gesture is produced. For the current study we use motion tracking of the silent gestures, here indicated with a pose-skeleton overlaying the original video data.

The two semantic dimensions provide possible axes for compression and systematization of the communicative tokens. At one extreme, one might invent 24 unique gestural utterances that are not clearly related to each other, such as in the following videos for “to sing” (<https://osf.io/d8srx/>) and “singer” (<https://osf.io/974ke/>). A more efficient encoding would be to start differentiating by functional category such that “microphone” is preceded by a general object marking gesture (“<https://osf.io/r3gcp/>”) and “singer” is preceded by a general person marking gesture (“<https://osf.io/ex4tv/>”), both followed by the same thematic marking gesture conveying “music”. Such general functional markers aid the disambiguation of related meanings, and they allow for a systematic reemployment of the same signal, thereby reducing the signaling space. In theory, once you invent 4 functional marker gestures, and 6 thematic marker gestures, you can systematically recombine these to convey all 24 meanings. The communicative system then has compressed its information density from 24 information units to 10 information units.  
Motamedi and colleagues (2019) indeed observed such signs of compression of the meaning space as the system developed. In early iterations of learning, fairly fine-grained iconic enactments were the most common way of gesturally depicting the referents. But over generations, functional markers were found to become more prevalent and meaning components were increasingly reused across gestures. On our supplemental page we have provided video examples (“<https://osf.io/5zqnb/>”) of all the gestures produced in generation 1 versus generation 5 for a particular chain (chain 1), where we highlight in red how in generation 5 there are clear recurring functional markers for the person category (using a pointing-to-self gesture) which is absent in generation 1. With meticulous hand coding of the different referential components of each silent gesture, Motamedi and colleagues (2019) quantitatively tested whether there was indeed *systematicity* emerging. The gesture coding included information about the form of a particular gesture segment, such as the number of manual articulators used (1 or 2 hands), as well as the referential target of the gesture (e.g., hat; pan; turn page). Based on the full sequences of the referential components that were uniquely expressed in each gesture, entropy was computed, which expresses compressibility of the gesture content, i.e., the amount of information that is needed to compress the signal set.  
When a lot of referential components in a gesture utterance recur between other gestures, such as in our theoretical case mentioned above, the system has a simpler structure and will show reuse of gestural components (e.g., Gibson et al., 2019). Dovetailing with the qualitative observations and other studies in this field (e.g., Verhoef et al., 2016), it was indeed found that gesture-component entropy decreased over the generations. Furthermore, the gestures were explicitly coded for the amount of marking for the functional category, and results showed that such gestures indeed occurred more often at later generations. Finally, average gesture duration - as a measure of communicative efficiency - did not reliably change over the generations, which ran counter to previous research showing reduction in complexity over repeated gesture use (Gerwing & Bavelas, 2004; Holler & Wilkin, 2011), as well as predictions that more mature communication systems tend towards maximal efficiency (Gibson et al., 2019).  
These results obtained in the lab resonate with findings from homesign (e.g., Haviland, 2013) and emerging sign languages (Senghas, Kita, & Özyürek, 2004). For example, it has been shown that in the expression of motion events first generation signers of Nicaraguan sign language performed more holistic presentations of path and manner, while in following generations manner and path were segmented. Such segmentations afford novel combinatoriality and therefore increases generativity of a language. It expresses the meaning space with fewer individual components, similar to how the participants studied by Motamedi et al. (2019) started to compress the meaning space by developing ways mark functional status across referents (e.g., “agent”, “action”).  
As we see here, qualitative empirical grounding is crucial for ensuring a rich understanding of evolving multimodal signaling systems. In addition, manual gesture coding makes it possible to track and quantify gestural elements encoding referents and functional dimensions, and yields measures that can be used in quantifications of entropy and the emergence of structure over time. However, it does so at the cost of reducing rich multidimensional kinematic signals to discretized sequences of coded values from a limited set. And because of this, some aspects of the evolving systems remain outside of our reach. Can we approach these developing systems in a way that is more sensitive to their continuous and dynamic properties? Can we preserve observations of emerging structure and go beyond them to trace more fine-grained similarities over time and across generations? Might we even see how transmission chains develop their own gestural ‘dialects’ over time? These are the questions we aim to address using kinematic methods, from a theoretical perspective influenced by dynamical systems and biosemiotics.

# 1 Current study

Here we build on data from this recent iterated learning paradigm with silent gestures (Motamedi et al., 2019). With computer vision (Cao, Simon, Wei, & Sheikh, 2017) we obtained motion traces of manual- and head gestures (see e.g., Lepic, Börstell, Belsitzman, & Sandler, 2016; Ripperda, Drijvers, & Holler, 2020). We then investigated kinematic interrelationships between gestures (e.g., Beecks et al., 2016, 2015; Sato, Schouwstra, Flaherty, & Kirby, 2020), where we leverage bivariate time series analysis (Dynamic Time Warping) with network topology analysis and visualization (Pouw & Dixon, 2019; Pouw et al., 2021). Through this analysis we show that the study of gesture’s form can reveal the linguistic constraints of the kinematic system as a whole.

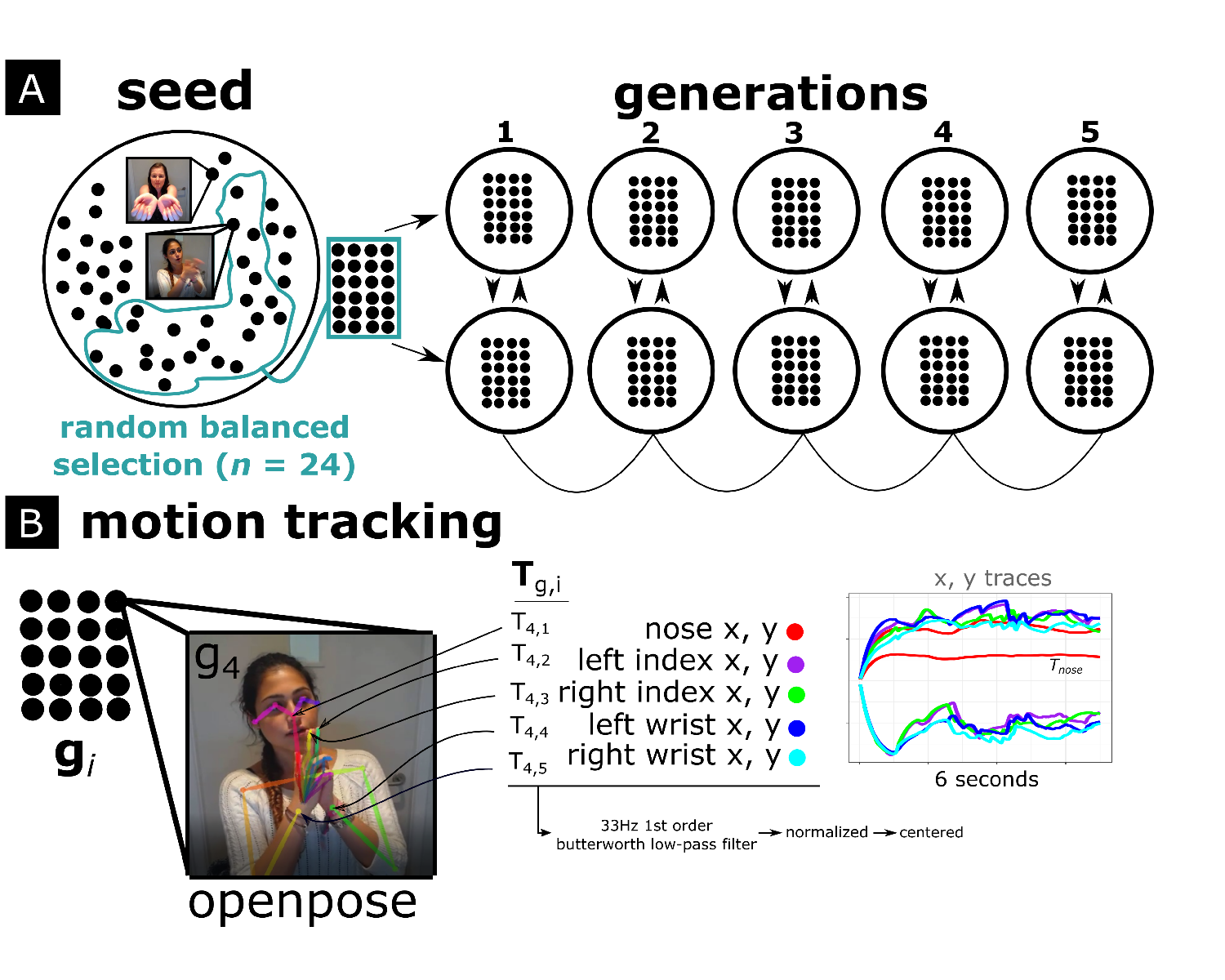
We hypothesized that over iterations:

* 1. gesture kinematics become simpler.
  2. gesture kinematic relationships become more systematic, and this scales with systematicity computed over gesture content from the original study.
  3. the simplification of kinematics at the level of individual gestures is related to the systematicity of kinematic relationships across gestures.
  4. idiosyncratic gesture cultures emerge as evidenced chains drifting away from each other over time.

Importantly, the prediction that gestures simplify is based on previous research reporting simplification as judged by human annotators (e.g., Gerwing & Bavelas, 2004) and earlier kinematic findings (Namboodiripad, Lenzen, Lepic, & Verhoef, 2016). We extend this research, as well as the original study (Motamedi et al., 2019), with a detailed kinematic analysis of the simplification of evolving gestures, assessing how not only salience (Namboodiripad et al., 2016), but also segmentation and temporality of gestures may change and simplify as they become part of a system of expression.  
Not all kinematic changes observed will be evidence of linguistic constraints however. Simplification over time, as in 1), could result simply from effort minimization. But when we observe increased systematicity in the system as a whole (2) as related to kinematic simplification (3), we have a direct indication of a systematically structuring communication system. Therefore, we assess whether a Shannon-based Entropy measure computed on kinematic relationships show that the systems becomes more structured (i.e., compressible). Using measures that are similar to the original study, we can then also compare whether the entropy of kinematics is related to the entropy of human-coded semantic content of the gesture. If so, we have a good evidence that linguistic constraints can be objectively studied from systematic changes in gesture form. Note, again, this is not just a methodological exercise to replicate original findings with automated methods. If we can show some equivalence between form and content analysis, then they are not on a qualitatively different analytic plane. If we can show that systematicity emerges over iterations without leaving the domain of kinematics, it means that a gesture’s form is more transparent to linguistic functioning than is currently assumed.  
Finally, with 4), we show how we can study chain-specific cultural evolutionary trajectories, by assessing the extent to which the communicative chains drift away from each other. The path-dependence of cultural evolution means that chains can diverge from one another over time, resulting in kinematic dialects. To assess this, we use cluster performance measures to quantify whether gestures within a chain tend to become more kinematically similar to each other and more dissimilar to gestures of other chains. This analysis is an example of the unique affordances of quantitative kinematic measures and will enable us to study, for the first time, the emergence of kinematic dialects.

# 2 Method

We will follow a bottom-up approach to the study of kinematics as communicative systems. In the first stage of our analysis, we demonstrate the specific changes that occur in the kinematics of the gestures. In a second stage, we assess possible systematic interrelationships in kinematic patterns of gestures through gesture network analysis (Pouw & Dixon, 2019). We will discuss each step in this procedure in the following sections, and finally discuss our main gesture network entropy measure. In the supplemental information we reserve extra space for sanity checks of automated processing and graphical descriptive results of the key measures.

Figure 2. Design experiment and openpose tracking  
 *Note Figure 2*. First steps of the general procedure. A) shows the original experiment setup (Motamedi et al., 2019), where a seed set of 24 gestures was randomly selected for each chain containing five generations. Seed gestures were used to train the first generation of each chain; after that, gestures from the previous generation were used as training data. Participants then communicated gesturally about the same concepts. B) For our analysis we first performed video-based motion tracking with OpenPose (Cao et al., 2017) to extract relevant 2D movement traces () for each gesture *g* for body key points *i* (nose, the wrists and index fingers). After motion tracking the next steps were dynamic time warping and gesture network analysis (Figure 3).

## 2.1 Participant, design, & procedure of the original study (experiment 1)

Here we discuss the setup of the experiment which generated the data we reanalyzed (for more detailed information see Motamedi et al., 2019).  
A seed gesture set was created with 48 pre-study participants who each depicted 1 out of 24 concepts. Thus for each concept there were two seed gestures performed by unique pre-study participants. Given that pre-study participants only produced one gesture, they were isolated from the other concepts that comprised the meaning space.  
For the main experiment (exp. 1) 50 right-handed English-speaking non-signing participants were recruited. They were allocated pairwise to one of 5 iteration chains. Participants were first shown a balanced subset of 24 unique seed gestures. These chain-specific seed gesture sets will be referred to as generation 0, which were followed by generations 1 through 5. In the training phase, gestures were presented in random order and participants were asked to identify the meaning of the gesture from the 24-item meaning space, followed by feedback about their performance. They were then asked to self-record their own copy of the gesture. Participants trained with a subset of 18 items (out of 24), and completed two rounds of training.  
In the testing phase, participants took turns as director and matcher to gesturally communicate (without using speech) and interpret items in the meaning space, with feedback following each trial. This director-matcher routine was repeated until both participants communicated all 24 meanings. Subsequent generations were initiated with new dyads whose training set was the gestures from one randomly selected participant from the prior generation.  
The recorded videos of the seed gestures and the gesture utterances participants produced in the testing phases are the data we use here. This means that we have 50 participants conveying 24 concepts = 1200 gesture videos belonging to generations 1-5, and 48 seed gesture videos with each concept conveyed by two unique seed participants. This forms the primary data that we reanalyse using kinematic methods.

## 2.2 Motion tracking

Motion tracking was performed on each video recording with a sampling rate of 30Hz. To extract movement traces, we used OpenPose (Cao et al., 2017), which is a pre-trained deep neural network approach for estimating human poses from video data (for a tutorial see Pouw & Trujillo, 2019). We selected keypoints that were most likely to cover the gross variability in gestural utterances: positional x (horizontal) and y (vertical) movement traces belonging to left- and right index fingers and wrists, as well as the nose. For all position traces and their derivatives, we applied a 1st order 30Hz low-pass Butterworth filter to smooth out high-frequency jitters having to do with sampling noise. We z-normalized and mean-centered position traces for each video to ensure that differences between subjects (e.g., body size) and within-subject differences in camera position at the start of the recording were inconsequential for our measurements.

## 2.3 Kinematic Properties

We first selected five potential measures representative of kinematic quality of the movements in terms of segmentation, salience and temporality, namely submovements, intermittency, gesture space, rhythm, and temporal variability (or rhythmicity). See Figure 3 for two example time series from which most measures can be computed. All measures were computed for each key point’s time series separately and then averaged so as to get an overall score for the multimodal utterance as a whole. Based on these exploratory measures we eventually selected three measures tracking gtracking gesture salience (gesture space), gesture segmentation (intermittency), and gesture temporality (temporal variability). We discuss the motivations for selecting each measure below. Correlations between these variables and distributions are shown in supplementary materials, Figure S1.

## 2.4 Gesture salience

As a measure for gesture salience or reduction, we computed a gesture space measure. This was determined by extracting the maximum vertical amplitude of a key point multiplied by the maximum horizontal amplitude, i.e., the area in pixels that has been maximally covered by the movement.

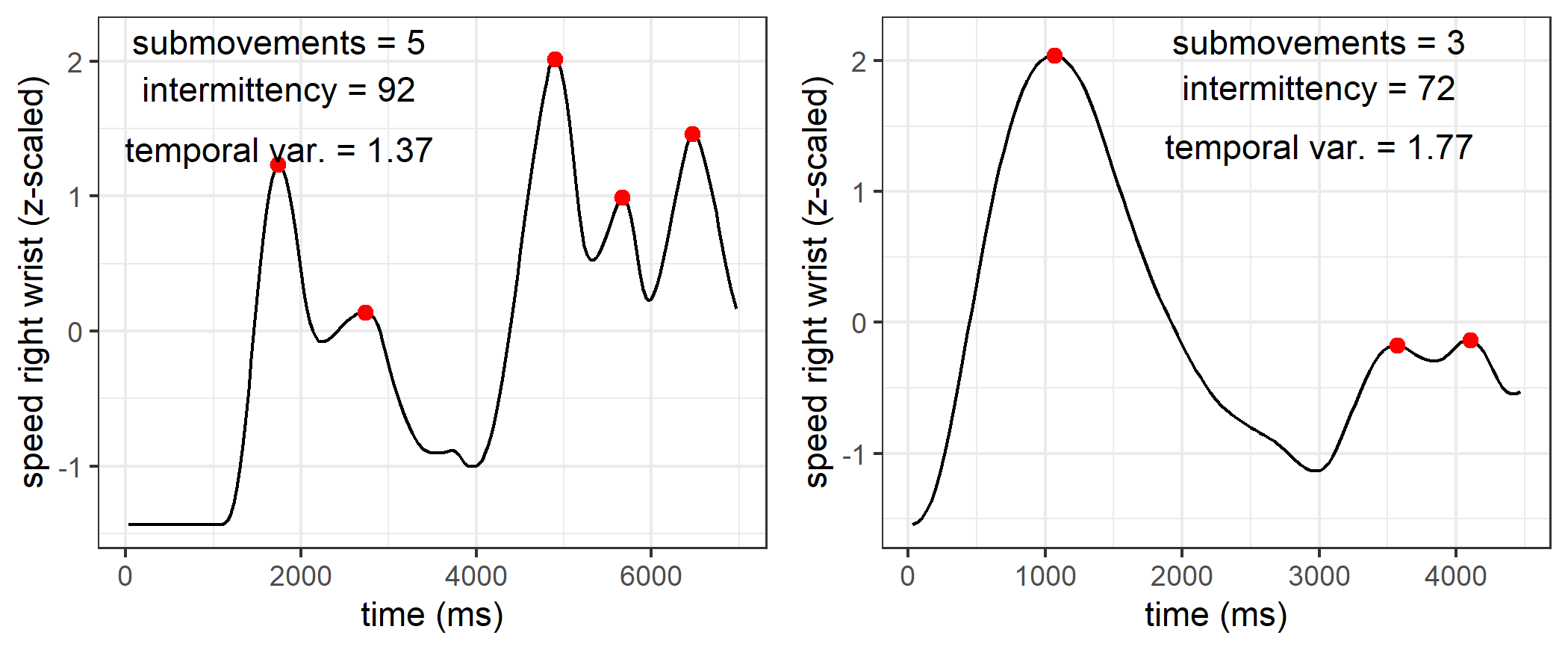
## 2.5 Gesture segmentation

We first computed a submovement measurement similarly implemented by Trujillo, Vaitonyte, Simanova, and Özyürek (2019). Submovements are computed with a basic peak finding function which identifies and counts maxima peaks in the movement speed time series. We set the minimum inter-peak distance at 8 frames, and minimum height = -1 (z-scaled; 1 std.), minimum rise = 0.1 (z-scaled).  
One property of the submovement measure is that it discretizes continuous information and uses arbitrary thresholds for what counts as a submovement, thereby risking information loss about subtle intermittencies in the movement. To have a more continuous measure of intermittency (the opposite of smoothness) of the movement we computed a dimensionless jerk measure (Hogan & Sternad, 2009). This measure is dimensionless in the sense that it is scaled by the maximum observed movement speed and duration of the movement. Dimensionless jerk is computed using the following formula:

Here is jerk (second derivative of the speed), which is squared and integrated over time and multiplied by duration cubed over the maximum squared velocity . We show in the supplementary materials (see supplemental figure S2) that this measure correlates very highly with submovements, thus we chose to only use intermittency for further analysis. Note that a *higher* intermittency score indicates more intermittent (less smooth) movement. We log transformed our smoothness measures due to skewed distributions.

### 2.5.1 Gesture temporality.

From the submovement measure we computed the average interval between each submovement (in Hz), which is a measure of rhythm tempo. This measure was, as expected, highly correlated with intermittency score (see S2), as tempo goes up when more segmented movements are performed in the same time window, *r* = 0.8, *p* = < .001, which led us to drop this measure for our analysis. Instead, we use another temporal measure that is more orthogonal to intermittency and gesture space, and which captures the stability of the rhythm; the temporal variability of the movements. This measure is simply the standard deviation of the temporal interval between submovements (given in Hz): a higher score indicates more temporal variability and a lower score indicates more isochronous rhythm. Note that this measure cannot be calculated when there are less than 3 submovements (i.e., when there no intervals in which we can detect the temporal variability).

Figure 3. Overview kinematic measures  
 *Note Figure 3*. Two time series showing right-hand wrist speed in two different trials. Our measures of segmentation and temporal variability are computed from time series like this. SEGMENTATION captures the amount of submovements (observed peaks in red), so the first time series is more segmented than the second. INTERMITTENCY captures similar information in a continuous fashion using rates of change in acceleration, yielding a higher score for the first time series than for the second. TEMPORAL VARIABILITY captures the rhythmicity of the signal, and is operationalized in terms of the regularity of temporal intervals between submovements. In the first plot, red dots occur at relatively equal temporal intervals (lower temporal variability), whereas in the second, the temporal intervals are highly unequal (higher temporal variability. Finally, gesture space was calculated from the size of x,y position traces not shown here.

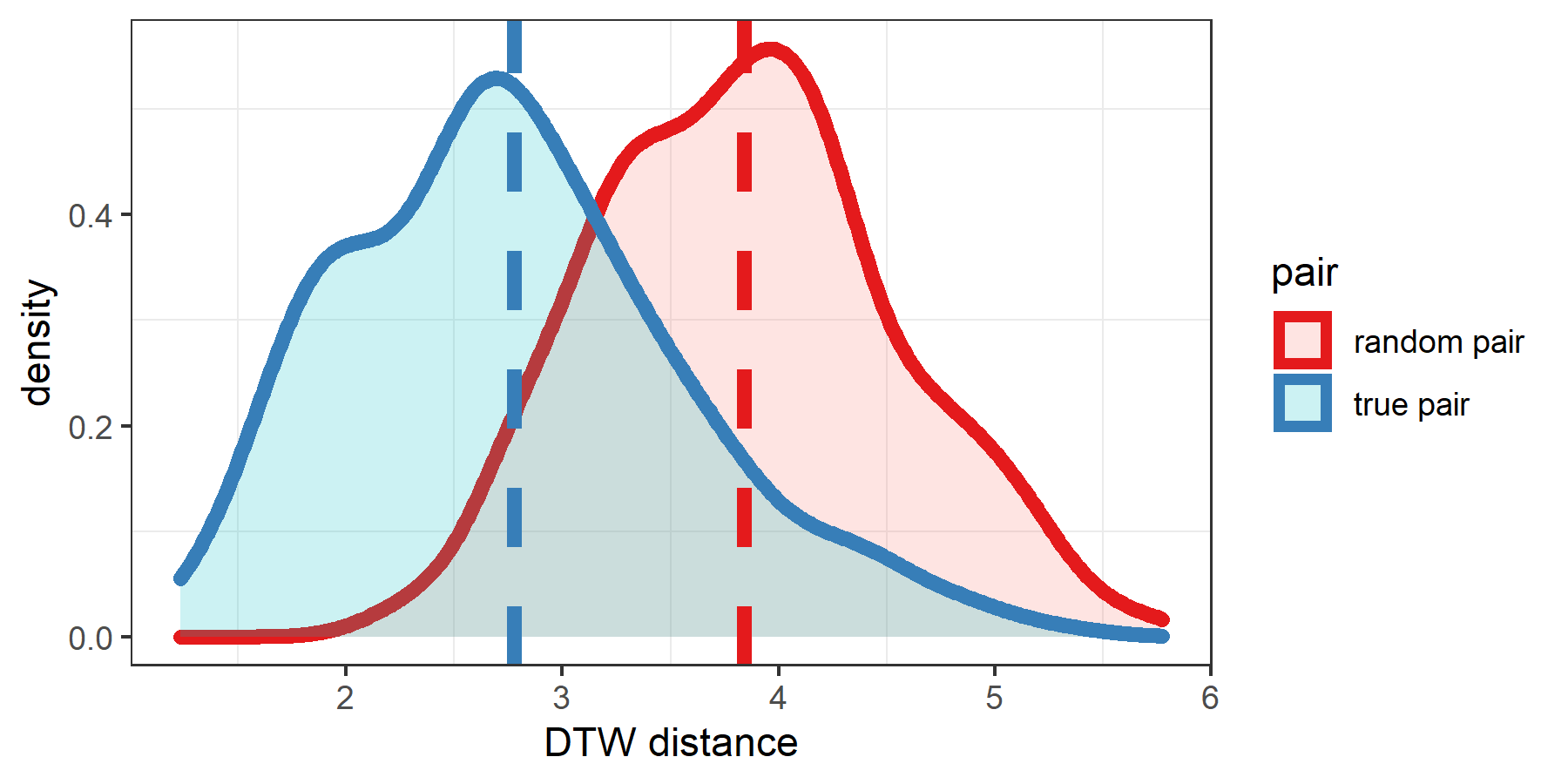
## 2.6 Human coding and kinematic measures

For information about how these automated kinematic measures approximate hand-coded data from Motamedi and colleagues (2019), see supplemental figure S2. The hand-coded data consisted of the amount of unique information units of the gesture utterance, the number of repetitions in the utterance, as well as the number of segments (information units + repetitions). We should predict that our kinematic intermittency score should correlate with the number of segments, repetitions and information units as the kinematics will have to carry those information units by contrasts in the trajectories. Supplemental Figure S2 shows the correlations for our kinematic measures and the human-coded gesture information. It shows that the amount of information units (unique, repeated, or total) in the gesture as interpreted by a human coder reliably correlate with kinematic intermittency (more intermittent, more human-coded information units), gesture space (larger space, more information units) and temporal variability (more stable rhythm, more information units).

## 2.7 Dynamic Time Warping

Dynamic Time Warping (DTW) is a common signal processing algorithm to quantify similarity between temporally ordered signals (Giorgino, 2009; Mueen & Keogh, 2016; Muller, 2007). The algorithm performs a matching procedure between two time series by maximally realigning (warping) nearest values in time while preserving order, and comparing their relative distances after this non-linear alignment procedure. The degree that the two time series diverge after warping indicates how dissimilar they are. This dissimilarity is expressed with the DTW distance measure, with a higher distance score for more dissimilar time series and a lower score for more similar time series.  
The time series in the current instance are multivariate, as we have a horizontal (x) and vertical (y) positional time-series data. However, DTW is easily generalizable to multivariate data, and can compute its distances in a multidimensional space, yielding a multivariate dependent variant of DTW. We opt for a dependent DTW procedure here as x and y positional data are part of a single position coordinate in space. Additionally, we have 6 of these 2-dimensional time series for each body key point. To compute a single distance measure between gestures, we computed for each gesture comparison a multivariate dependent DTW Distance measure per key point, which was then summed for all keypoint comparisons to obtain a single Distance measure D (illustrated in Figure 5). The D measure thus reflects a general dissimilarity (higher D) or similarity (lower D) of the whole manual+head movement utterance versus another utterance. Note the DTW procedure is applied on the entire gesture utterance, which could consist of multiple components (e.g., hand cuff gesture + pointing).  
We used the R package ‘DTW’ (Giorgino, 2009) to produce the multivariate distances per keypoint. The DTW distance measure was normalized for both time series’ length, such that average distances are expressed per unit time, rather than summing distances over time which would yield higher (and biased) distance estimates for longer time series (i.e., longer gesture videos). For further conceptual overview and methodological considerations of our DTW procedure see Pouw and Dixon (2019).

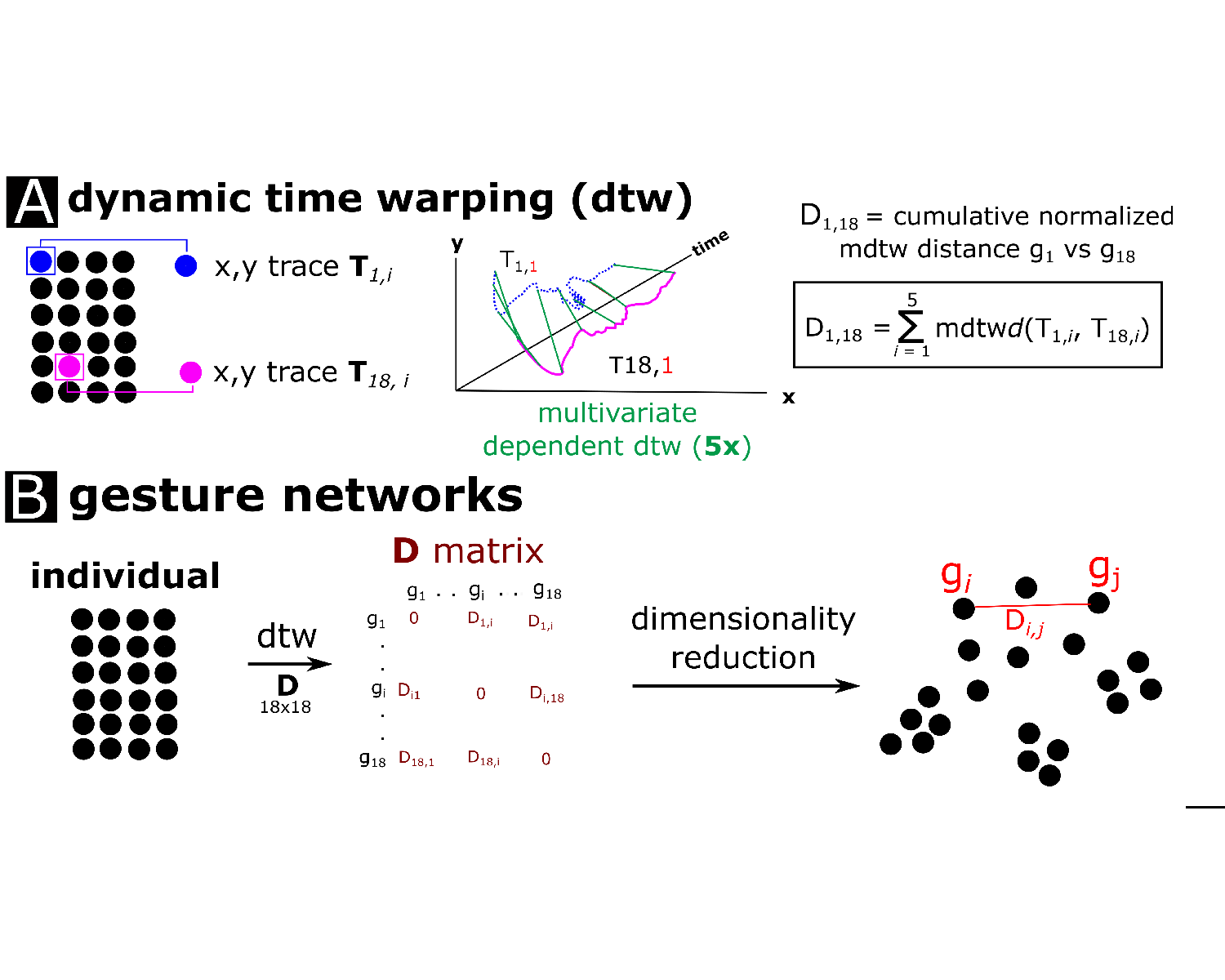
As a demonstration that our D measure reflects actual differences in kinematics, we computed for each individual in each chain the difference between a gesture seed and the gesture that the individual produced to copy it, for generation 1. These “true pairs” must be maximally similar (lower D) as the individual produced their copied gesture shortly after first exposure in the training phase, which should lead to high faithfulness in reproduction. We contrast this with a false or random comparison of the same gesture in generation 1 with a gesture seed that was neither in the same functional nor thematic category. These false random pairs must be more dissimilar, and should produce higher DTW distances. Figure 3 shows the distributions of the distances observed. DTW distance distributions were reliably different, *t* (469.77) = 15.82, *p* = < .001, Cohen’s *d* = 1.44, for the true pair, *M* =2.78(*SD* = 0.78), as compared to the random pair, *M* = 3.84(*SD* = 0.69).  
We find that adding head movement trajectory to our D calculation significantly increases false-real pair discriminability as compared when we compute our D measure on only manual keypoints (left/right wrist and index fingers), change in Cohen’s *d* = 0.41, change D real vs. false = 0.33, *p* = < .001. Therefore, we conclude that in the current experiment the gesture utterances are also crucially defined by head movements as well. This is a novel finding in and of itself, and demonstrates the multi-articulatory nature of silent gestures.

Figure 4. Density distributions of D for true pairs and random pairs  


*Note Figure 4*. Density distributions of D are shown for the random versus real pairs. With D based on head-, wrist- and finger movement there is good discriminability between real versus falsely paired gestures, confirming that our approach is tracking gesture similarity well.

## 2.8 Gesture kinematic networks

Graphically shown in Figure 5D, we constructed for each participant (nested in generation and chain), as well as each seed gesture set (seed set belonging to that chain), a distance matrix **D**, containing the continuous D comparisons for each gesture produced by that participant with each other gesture produced by that participant, yielding a 24x24 matrix. The diagonal contains zeros for gesture comparisons that are identical (). These characteristics make **D** a weighted symmetric distance matrix.  
For each distance matrix we can construct a visual geometric representation of its topology by projecting the distance of gesture tokens on a 2d plane using a dimensionality reduction technique called “t-SNE”, a variant of Stochastic Neighbor Embedding (Maaten & Hinton, 2008). These 2-d representations show locations of gesture nodes, with distances between gesture nodes approximating our D measure. Such 2D representations are imperfect approximations of the underlying multidimensional data and are only used as visual aids. The uncompressed distance matrices are used to calculate entropy and other measures. We refer to these measurements as ‘network properties’ as these measures are intuitively understood in network or geometric terms. For calculations of network entropy we use the R package ‘igraph’ (Csárdi, 2019), and for dimensionality reduction we use R package ‘tsne’ (Donaldson, 2016). On our supplemental page we again show video examples of all the gestures produced in generation 1 versus generation 5 for a particular chain (chain 1), but now with videos spatially located according to their coordinates in kinematic space (<https://osf.io/wbmf9/>). Examples are highlighted in red, where kinematic similarity increases from generation 1 to 5 due to functional markers being used for the category “location”.

Figure 5. General method gesture network analysis  *Note Figure 5*. C) For each gesture comparison within a gesture set, the time series were then submitted in to a Dynamic Time Warping procedure where we computed for each body part a multivariate normalized distance measure, repeated for all body parts and summed, resulting in one overall distance measure D for each gesture comparison. D) All distance measures were saved into a matrix **D** containing all gesture comparisons within the comparison set, resulting in a 24x24 distance matrix. The distance matrix can be visualized as a fully connected weighted graph through dimensionality reduction techniques, such that nodes indicate gesture utterances and the distance (or weight) between gesture nodes representing the ‘D’ measure, indicating dissimilarity.

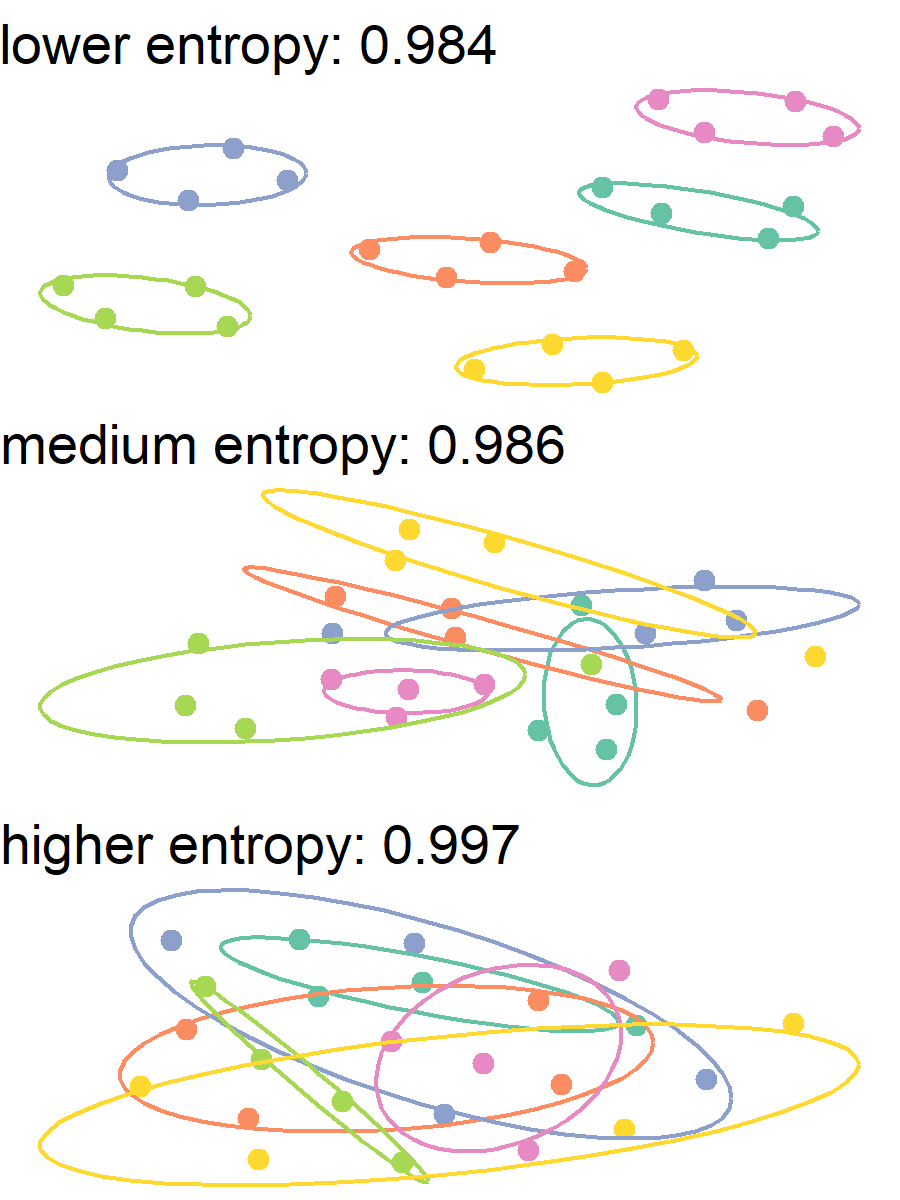
### 2.8.1 Kinematic Entropy.

Entropy is a measure that quantifies the compressibility of data structures, and has been used to gauge the combinatorial structure of communicative tokens in the field of language evolution (e.g., Verhoef et al., 2016; for theoretical grounding see Gibson et al., 2019). In the original experiment, Motamedi and colleagues (2019) computed entropy from the gesture content codings, which captured recurrent information units between gestures. In our case, entropy quantifies the degree to which there are similar or more diverse edge lengths (i.e., similar/diverse levels of dissimilarity ‘D’ between combinations of two gesture trajectories). If they are more similar, lower entropy reflects that communicative tokens relate to each other in more structural ways. So our measure of network entropy gauges how compressible kinematic interrelationships are, which is conceptually related to the systematic recurrence of information units between the human judged gesture content. The network entropy measure we used (see Eagle, Macy, & Claxton, 2010) is almost identical to a classic Shannon entropy calculation used in the original study to quantify the systematicity of the gesture’s content (Motamedi et al., 2019), where . The only difference is that our measure is computed on the distances for each node relative to the shortest path to the other nodes, and then normalized by the number of gesture distances. So our measure quantifies the topological diversity of the gesture relationships, where a lower score indicates more similar relationships and a higher score indicates a more randomly distributed set of relationships. Specifically, for each gesture node we compute the diversity of kinematic distances to other gestures, using a scaled Shannon Entropy measure:

Here, is the number of gesture connections for gesture , and is the proportional distance:

Here is the distance between gesture and gesture divided by the total distance involving gesture . Figure 6 shows a graphical example of different network structures and the concomittant entropy measure.

Figure 6. Example network entropy



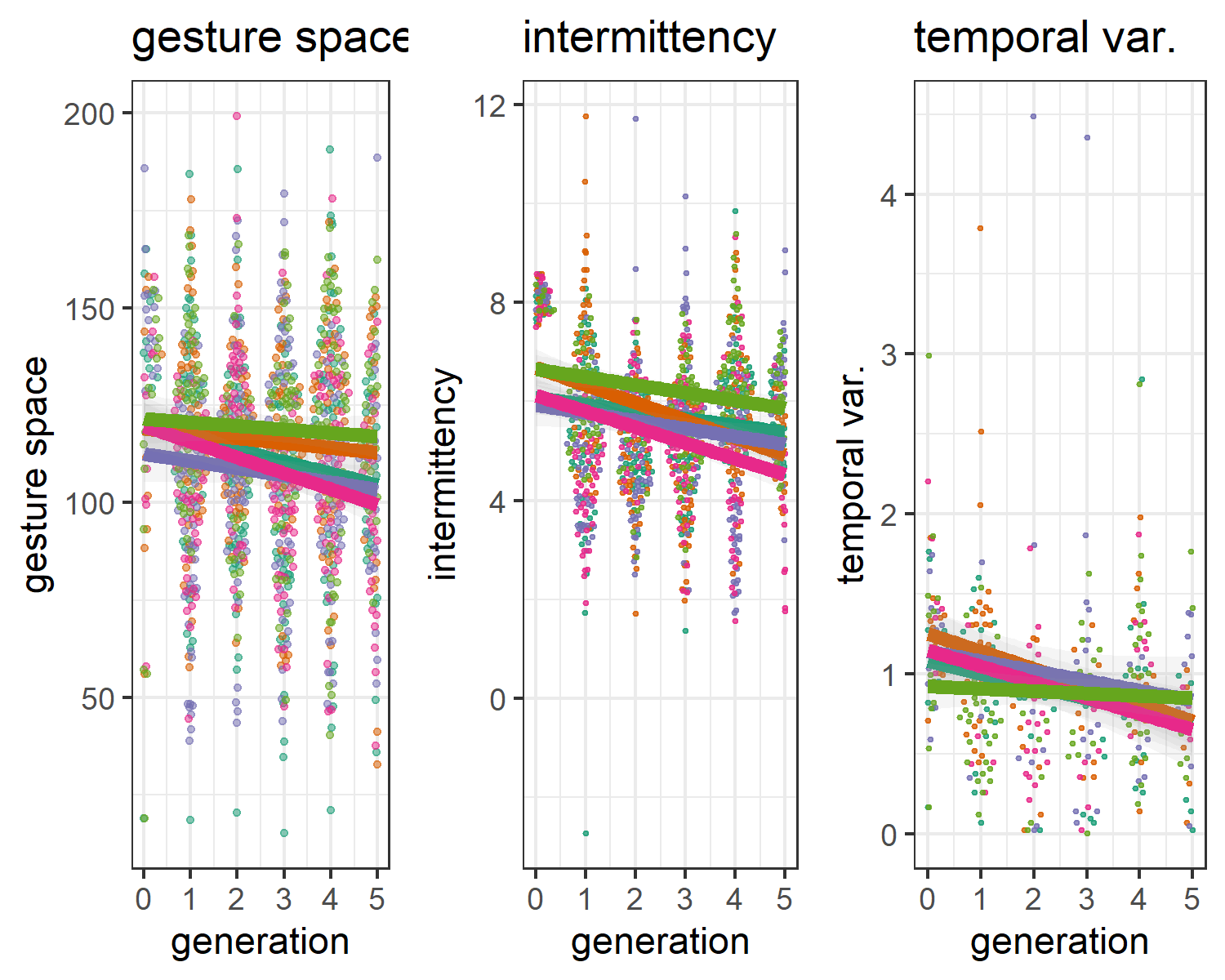
*Note Figure 6*. Simulated data showing 6 clusters with low variance in distances (top panel), higher variance (middle panel), or randomly distributed distances (lower panel). More variable and random distributions of node distances yield higher entropy scores. In contrast, entropy is lowest when interrelationships are distributed in a more systematic way (top panel).

# 3 Main Results

We first report changes in kinematic features over generations. Then we consider change in relations between communicative tokens over generations, as indexed by kinematic network entropy. We also relate kinematic changes to network-level changes. Finally, we consider how chains diverge over time, allowing a peek into the emergence of unique gesture cultures.

## 3.1 Kinematic features

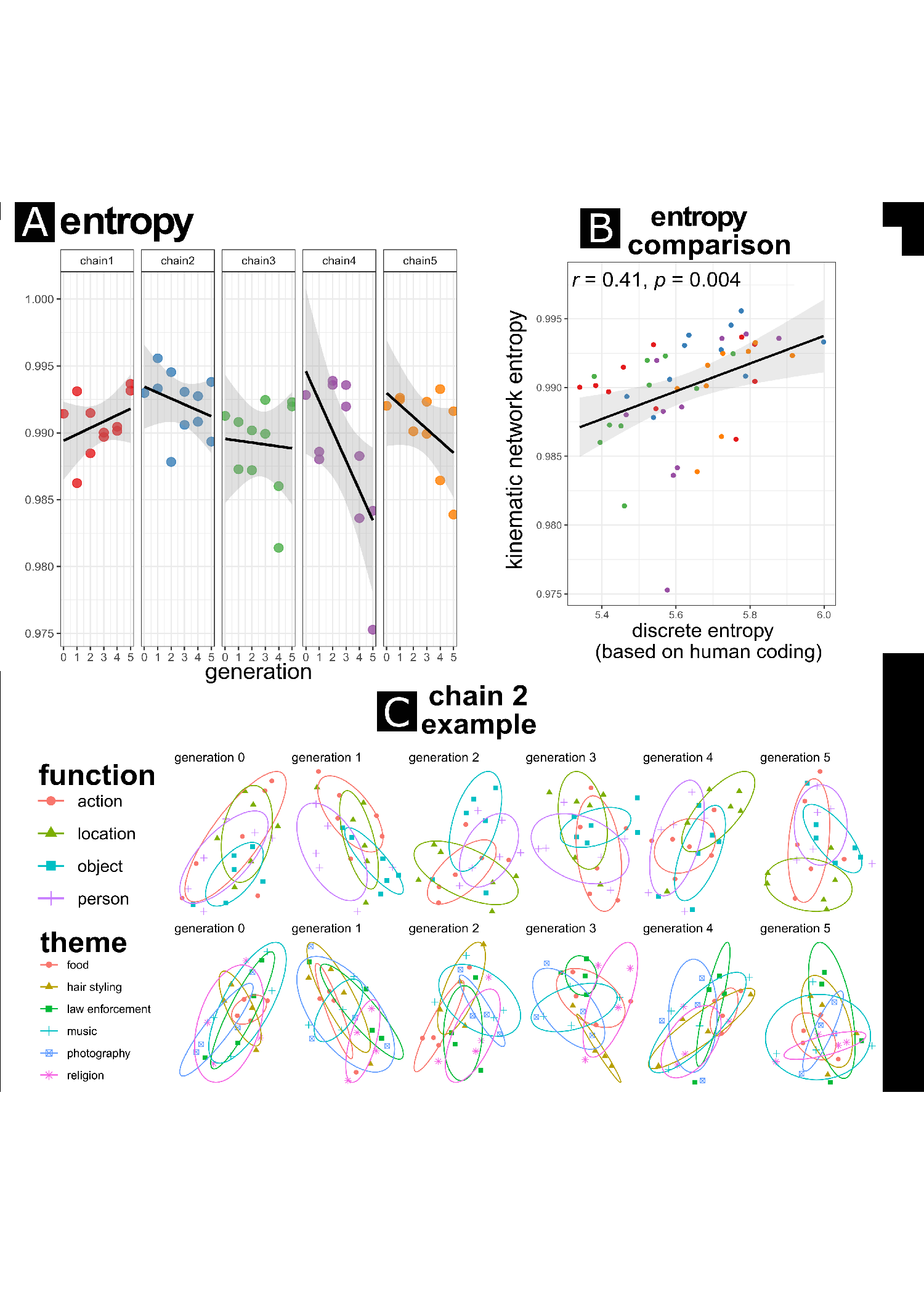
A key aim of our analysis is to capture the fine-grained kinematic features that drive changes in the gestural systems over generations, and which are hard to capture with a manual coding system focusing on the semiotic relation between gesture and meaning. All three of our kinematic measures show the hallmarks of increased communicative efficiency through reduced kinematic complexity over generations (Figure 7).

Figure 7. Change in kinematic properties over generations  
 *Note Figure 7*. Generation trends per chain for intermittency, temporal variability and gesture space. Over the generations, movements become more smooth (lower intermittency score), show more stable rhythms (lower temporal variability), and more minimized movements (smaller gesture space). There are fewer data points for temporal variability because this can only be computed for comparisons of gestures that have more than 2 submovements. So temporal variability indicates that *when there was a multi-segmented movement*, then such movements were more rhythmic.

We performed mixed effects regression analysis for assessing potential kinematic changes as a function of generation, with random intercept for participants nested within chains (random slopes did not converge). Generation reliably predicted intermittency of the movements relative to a base model, chi-squared change (1) = 76.66, *p* <.001, model *R*-squared = 0.06. In this model generation predicted lower intermittency score, *b* estimate = -0.2263, *t* (1,135.00) = -8.90, *p* < .001, Cohen’s *d* = -0.53). We also observe lower temporal variability as a function of generations, chi-squared change (1) = 24.12, *p* <.001, model *R*-squared = 0.05, indicating more stable rhythmic movements at later generations, *b* estimate = -0.0693, *t* (332.00 ) = -4.97, *p* < .001, Cohen’s *d* = -0.55)).  
Finally, over the generations gesture space decreased, chi-squared change (1) = 24.45, *p* <.001, model *R*-squared = 0.00. Model estimated gesture space was less for later generations, *b* estimate = -2.2100, *t* ( 1,174.00) = -4.97, *p* < .001, Cohen’s *d* = -0.29).  
Subtle changes in kinematic features are hard to capture using human coding, and indeed the rough proxies for this used by Motamedi and colleagues (2019) (length and number of repetitions of coded information units), did not demonstrate increased communicative efficiency. Here, we are able to capture increased efficiency by quantifying fine-grained kinematic features at the level of gesture tokens. Using independently motivated measures we found that gestures were on average smaller, less temporally variable, and less intermittent as the communicative system matured.

## 3.2 Network changes over generations

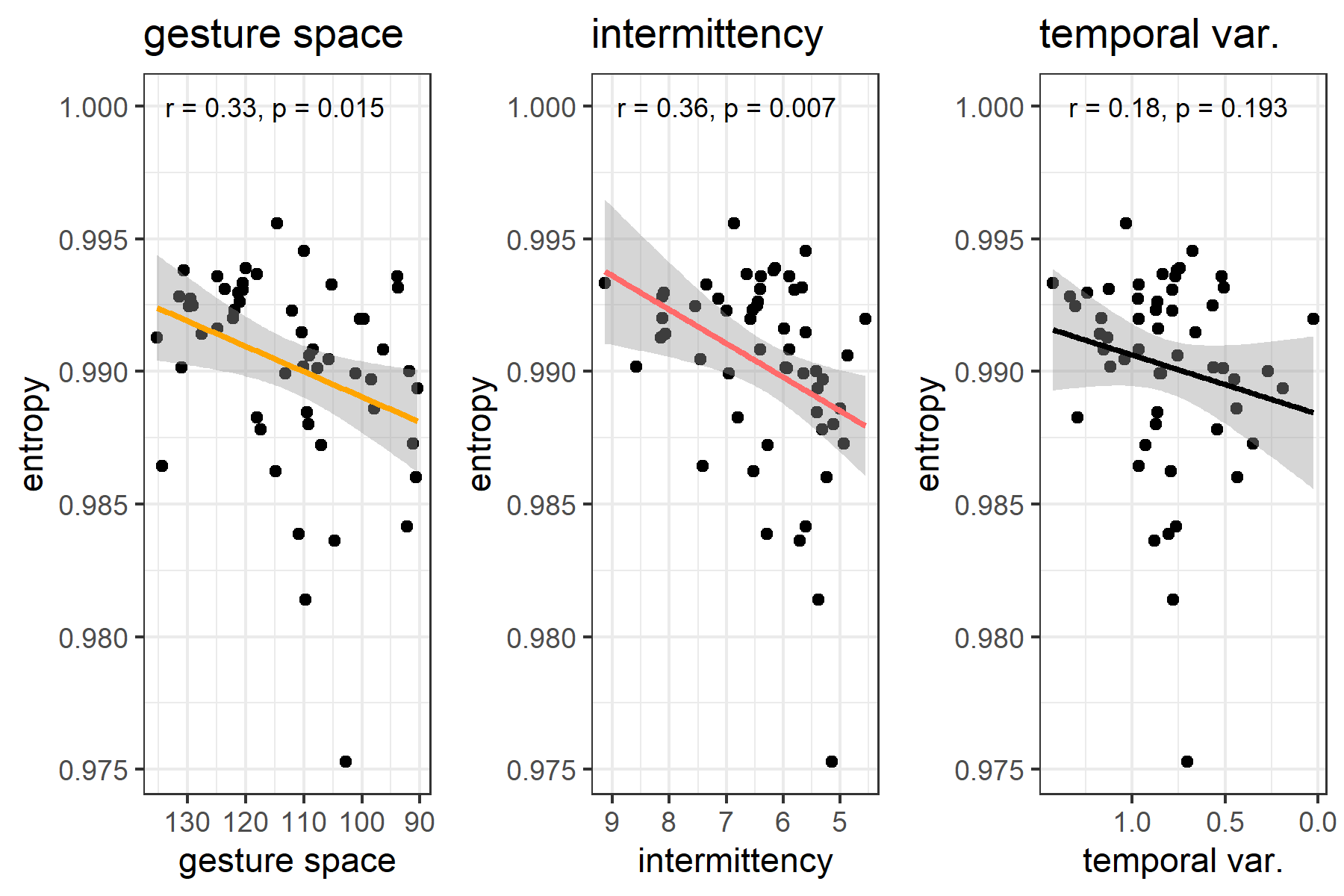
While changes in kinematic complexity suggest an increase in efficiency, they do not by themselves provide evidence of systematicity, another hallmark feature of communicative systems. Here we assess whether the gesture network as a system shows reduced entropy over generations, which would mean that the interrelationships between gestures become less randomly distributed. Figure 8a shows that the entropy of gesture networks indeed decreased as a function of generation in 4 out of 5 chains, indicating lower complexity of gesture interrelations as the systems matured. This reduction in entropy, it turns out, scales very reliably with the discrete entropy derived from the manual coding (Fig 8b).

Figure 8. Changes in networks measures over generations within chains  *Note Figure 8.* Panel ‘a’ shows for each chain the changes over generations in entropy, with generation 0 indicating the seed gesture set. For each generation > 0 there are two data points as there are two participants in each generation. Entropy tends to decline over the generations, indicating that relationships between tokens became less diverse, possibly indicating systematicity in the way nodes are connected. Panel ‘b’ shows that the network entropy computed on the kinematic distances was scaled reliably scaling with the discrete entropy computed on human annotated data, which suggests that our kinematic derived measure captures a comparable phenomenon in the evolution of gesture. In panel ‘c’ we provide the 2D network representations for each generation in chain 2, color coded and ellipses drawn for theme and function categories for separate rows. We arbitrarily picked one participant for each chain iteration. Note, it is difficult to directly see changes in structure in these representations which can be directly related to reduction in entropy, but in general the distances are less diverse over the generations.

We tested the trend indicated in Figure 8a in a mixed effects regression model similar to the original study (Motamedi et al., 2019), with chain as random intercept (random slopes did not converge for these models) and generation as independent predictor (0-5 generations, with generation 0 being the seed gesture network). Generation was indeed a reliable predictor for network entropy as compared to a base model predicting the overall mean, chi-squared change (1) = 4.75, *p* = 0.03, model *R*-squared = 0.08. Model estimates showed that entropy decreased over generation, *b* estimate = -0.0006, *t* (48.00) = -2.19, *p* = 0.03, Cohen’s *d* = -0.63).  
In sum, we find that kinematic network entropy decreases over generations, suggesting a steady increase in systematicity in terms of the distance between gestures in gesture networks. Furthermore, there is a clear scaling relation between the gesture-content entropy computed on human categorical codings and our measure of kinematic entropy, suggesting we are capturing a similar systematic property of these evolving manual languages.

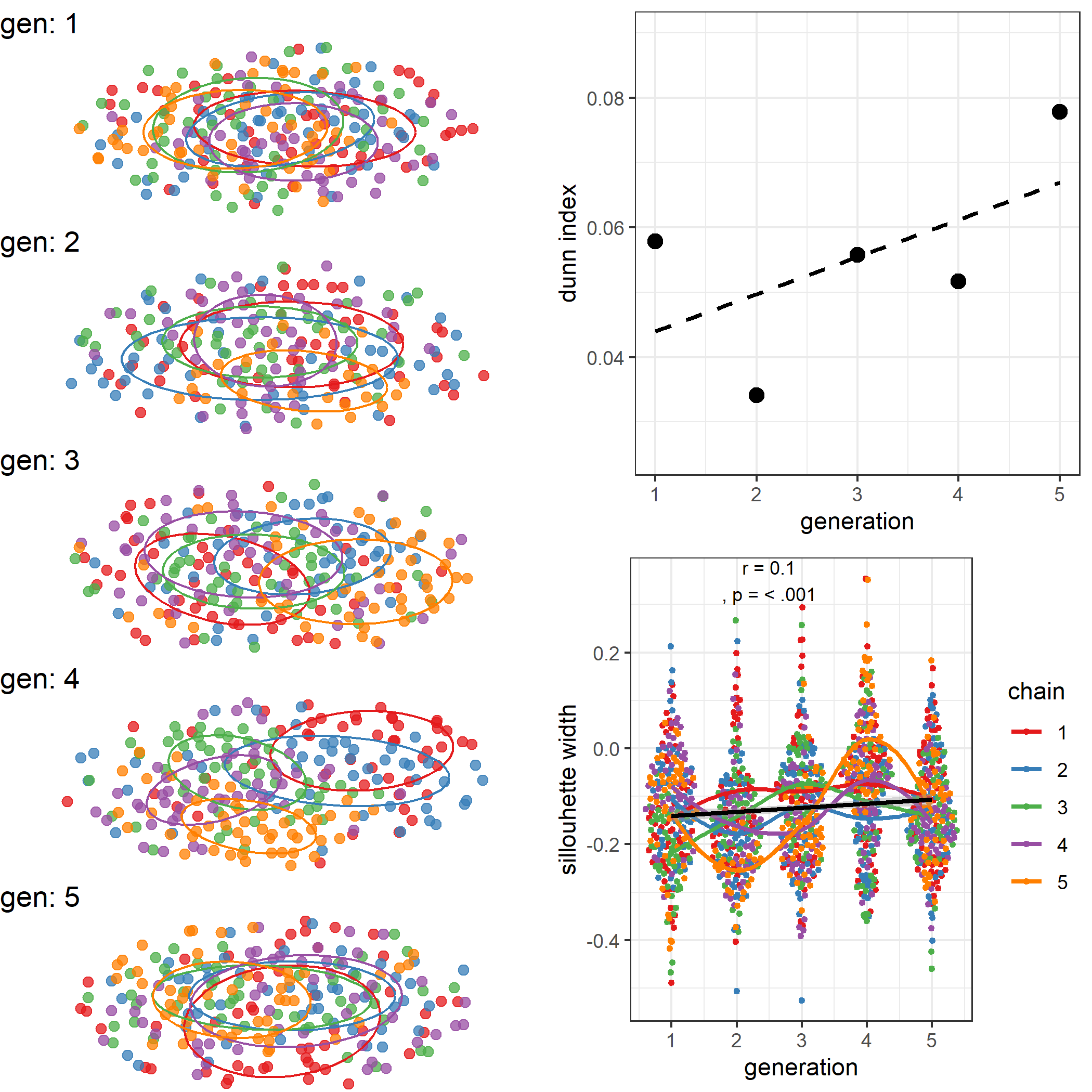
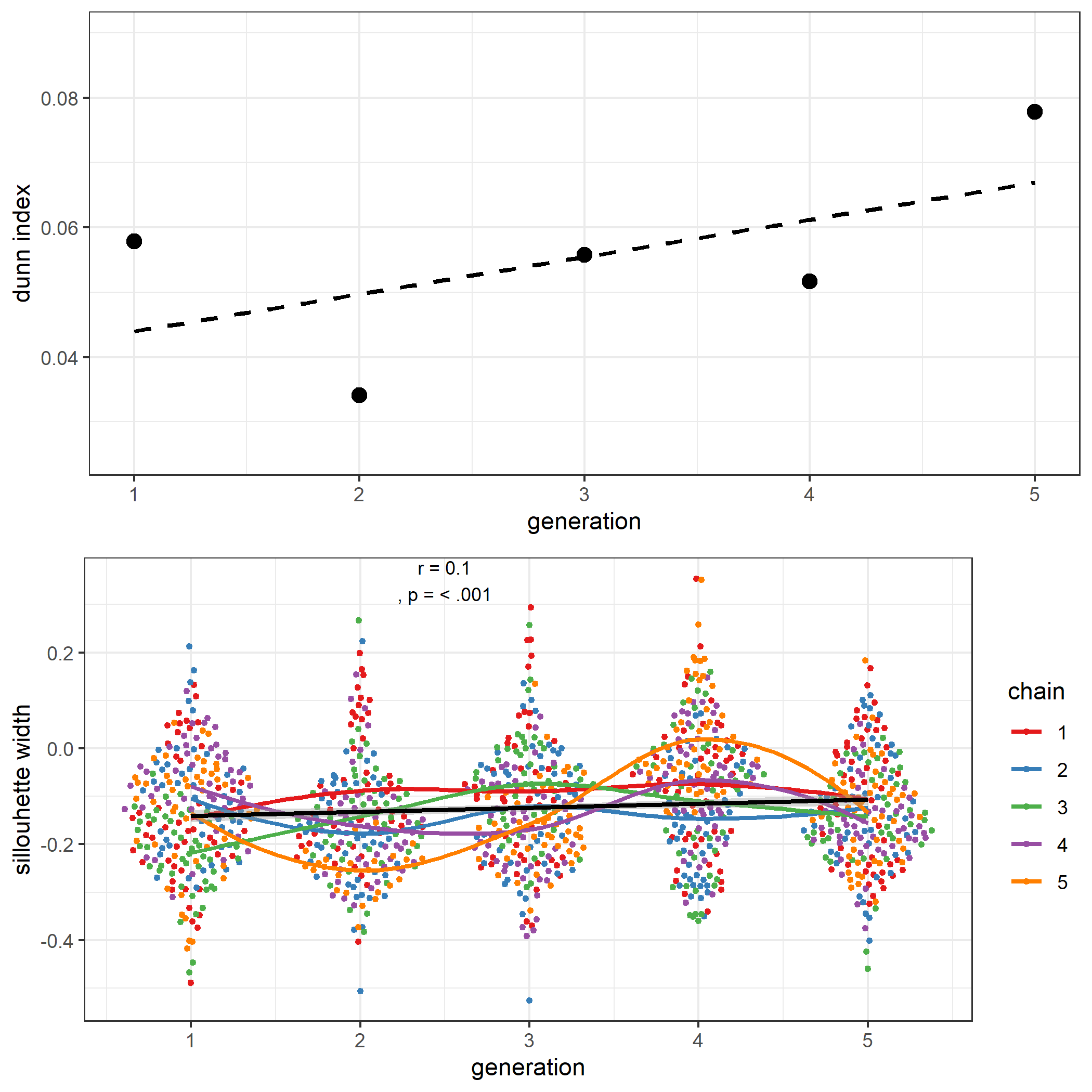
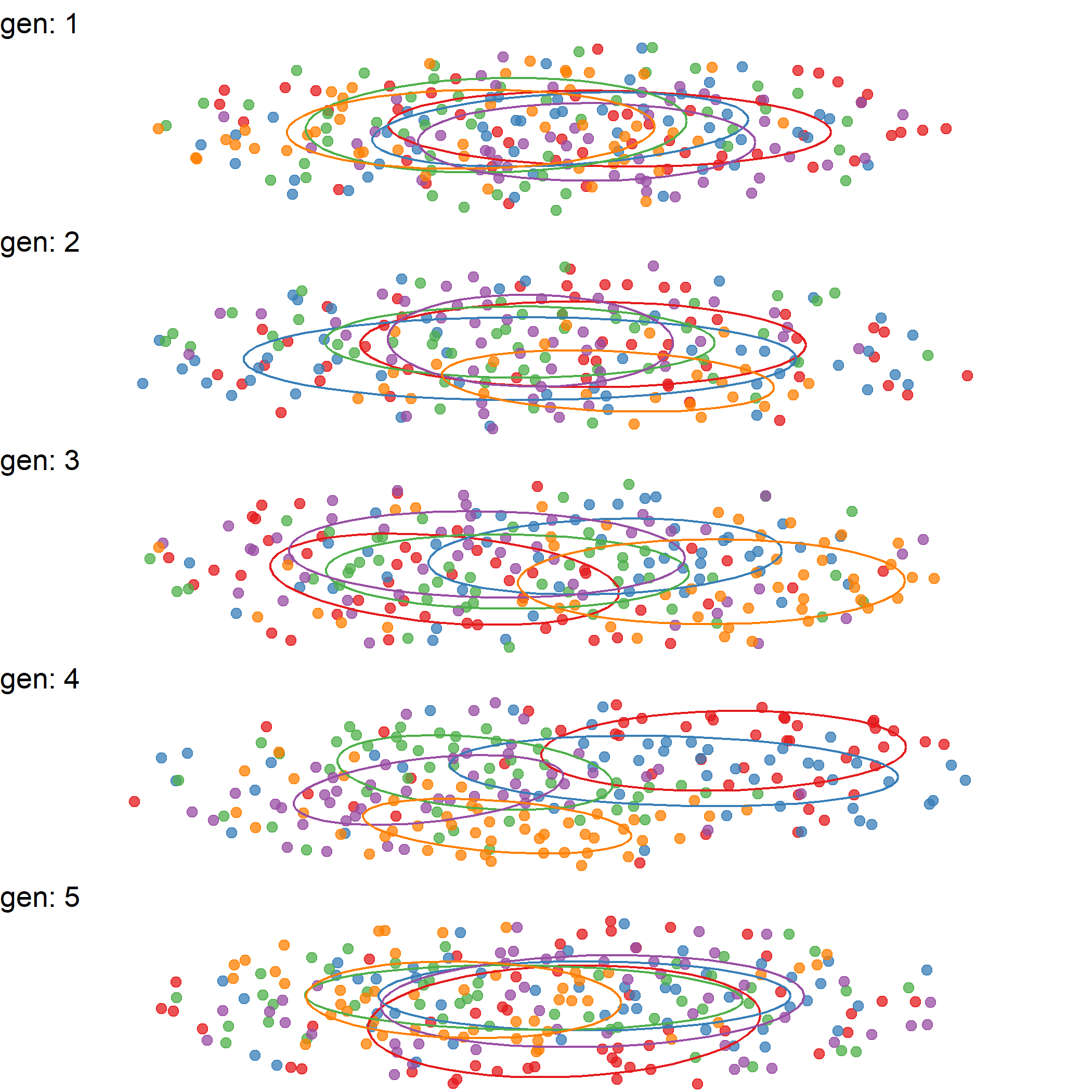
## 3.3 Relations between Kinematic and network properties

So far we have shown an overall increase in communicative efficiency (as measured by change in kinematic features over generations) and an increase in systematicy (as measured in decreasing entropy over generations). Reduction in kinematic features may or may not be related to the systematicity found in each gesture chain. Figure 9 shows how each kinematic property (averaged by participant) relates to gesture network entropy. We see that network entropy reduces as the average gesture space decreases and movements become less intermittent.

Figure 9. Relation between kinematic properties and network measures  *Note Figure 9*. Correlations are shown for each kinematic property averaged over all tokens for that participant and the concomitant network entropy for that participant. Note that the x-axis is reversed such that reductions in kinematic complexity are related to entropy. It can be seen that less intermittency and smaller gesture spaces, relate to lower entropy. This indicates that especially reduction in intermittency and gesture space is related with emergence of systematicity on the network level.

## 3.4 Chain-unique kinematic evolution over the generations

We have obtained evidence that over generations, kinematic changes go hand-in-hand with a decrease in complexity, and kinematic interrelationships between gestures indicate increasing systematicity. In our final analysis, we use our fine-grained token-level kinematic measures to assess whether unique trajectories in language evolution can be observed. The kinematic changes we have detected could suggest that motoric constraints lead iconic gesture systems to become more similar to each other, rather than as unique language-like systems. To test whether this is the case, we analyze whether the kinematic interrelationships within a chain diverge from other chains across the study. We predict that even though chains converge in the degree of systematicity (as we found above), they may actually diverge from each other over generations, as conventions are built that are unique to each chain. Figure 10 shows the main results of this research question.

Figure 10. Evolution of kinematic interrelationships between and within chains over the generations  
 *Note Figure 10.* (A) Network representations of kinematic interrelationships colour-coded by chain and with ellipses showing cluster centroids in similarity space. In generation 1, nodes are scattered and centroids mostly overlap, suggesting gestures do not clearly pattern by chain yet. Over generations, gestures increasingly cluster by chain, and chains drift apart, as seen in the decreasing centroid overlap in generations 3-5. Right panel: For gestures grouped by chains, the Dunn index shows an increase in cluster compactness and distinctiveness over generations (B), and Silhouette width reveals at the level of single gestures how the mean distance to tokens from other chains reliably grows relative to the mean distance to tokens from the same chain (C). This is for example apparent when contrasting chain 5 at generation 2 versus that same chain at generation 4, showing higher silhouette scores at generation 4 and more compact clustering in (A).

The network representations at each generation show that early on, gestures do not cluster clearly by chain. However, over generations, especially in generation 3 to 5, we see that gestures from particular chains start to cluster together more prominently, and clusters move away from each other to some extent, indicating greater chain-internal similarity and growing differences across chains (Figure 10A).  
Clustering can be quantified in several ways. Here we provide two well-known cluster performances measures: the Dunn index and the Silhouette width (Yadav, Tomar, & Agarwal, 2013). In general, cluster performance measures relate the within-cluster distances between nodes (which is minimal when clusters are stable), to between-cluster distances (which is maximal when clusters are stable), though they vary in how they compute the within and between distances. The Dunn index quantifies the compactness of the clusters assigned (chains in our case) and relates the minimum distance between centroids of each cluster to maximal distance between points, where higher values indicate better clustering.This measure shows a clear increase in chain-clustering performance over generations (Figure 10B). However, this measure only yields 5 data points in our case, one for each chain, which makes it hard to perform a statistical test. Therefore we also computed a token level measure of Silhouette width, which, for each token, relates the mean distance to other tokens within its cluster to the minimum distance between a member of a neighboring cluster. We find that, over generations, Silhouette width increases reliably, *r* = 0.1, *p* < 0.001 (Figure 10C).  
In sum, we find that each chain creatively shapes their own gesture system. While our network analysis showed that gesture systems become internally more coherent for all chains, this does not mean that chains become similar to each other. Instead, they forge their own developmental trajectory through over generations.

# 4 Discussion

Based on signal processing alone we have detected systematic changes reflective of a linguistically maturing communication system, from continuous multi-articulatory kinematics of silent gestures. We applied computer vision techniques to extract kinematics from video data (e.g., Östling, Börstell, & Courtaux, 2018; Ripperda et al., 2020; Trettenbrein & Zaccarella, 2021). We then quantified kinematic relationships between gestural utterances (Pouw & Dixon, 2019). Our analysis showed that a coding system focusing primarily on the hands would miss important gestural content that can be observed from head movements. We found that including head motion data as opposed to only motion data from the hands allowed us to better distinguish randomly paired gestures from directly related. In line with earlier research suggesting that communicative systems tend towards communicative efficiency (Gibson et al., 2019; Namboodiripad et al., 2016)), and going beyond the original study findings, we obtain that gestures reduce their complexity over the generations, by reducing size, submovements, and becoming more rhythmic. Communicative efficiency does not automatically entail systematicity. Therefore, we analyzed gesture networks and showed that communicative tokens have higher systematicity, a finding that is in line with measures derived from manual coding. These results suggest that form analysis can provide information about systematicity, making our method a valuable addition to the toolkit of gesture analysis (where it complements human coding of gesture content) and the cultural evolution of signaling systems more generally and without resorting to categorical semantic coding.  
The equivalence of form-based kinematic analysis and content-based manual coding is as much a theoretical advancement as it is a methodological one. It means that aspects of form can directly embody some of the systematic structure inherent in a system (Rączaszek-Leonardi & Kelso, 2008). This is underlined by our final, novel analysis, in which we found that while chains show a decrease in systematicity, the gestures that are produced become more distinct with respect to the other chains. Gestures within chains became clearly more clustered over time, meaning their form became more alike within the chain at later generations, and more dissimilar from other chains. This is clear evidence of a drift towards chain-specific conventions, suggesting the emergence of gesture kinematic dialects. In sum, the current analysis provides new insights in how gesture movements linguistically evolve. First, increased communicative efficiency was previously judged absent on the basis of gestural unit length, but we show it is present in the kinematics on relevant dimensions as salience, segmentation, and temporality. Second, kinematic properties change coherently to simplify in structure, which it turns out, is relatable to the simplification of the semantic content of gestures; a non-trivial equivalence given the fact that form and content are generally forcefully distinguished. Finally, we obtain that kinematic dialects emerge, which goes beyond the goals of the original study altogether.

## 4.1 Shortcomings and advantages

There are two important caveats to the analyses presented here. First, the kinematic analysis cannot say anything precise enough to determine the semiotic content of tokens (cf. Pouw et al., 2021). Therefore, understanding the semiotic content of human communication will always require extensive human analysis (Sandler, 2018). While understanding the semiotic content requires human analysis, kinematic analysis provide us with a unique grasp of aspects of rich evolving communicative systems that may elude human coders or would be too resource-demanding to manually code. The most productive way forward therefore is to use methodological triangulation by combining kinematic measures and semiotic analyses. For instance, kinematic measures would allow the detection of subtle changes in gesture form or system-level structure not possible through semantic analysis alone. This analysis can then be subjected to linguistic analysis or human coding to understand semiotic and structural aspects of the evolving system. For example, we find that the inclusion of head movements meaningfully improves our kinematic measures. While the original experiment focused on the hands as the most important articulators (Motamedi et al., 2019), our finding invites consideration of how head movements may be recruited as part of a culturally evolving semiotic system, a finding that has implications beyond purely methodological concerns. The general picture that emerges is that kinematic measures allow a bottom-up, data-driven approach to be informed and enriched by semiotic analysis at particularly quantitatively “active regions” of the data.  
A second caveat concerns a limitation to dynamic time warping. One finding from prior work on the evolution of compositionality is that holistic gestures may become segmented, freeing up individual elements to be recombined, possible in different orders. Now consider two gestures that contain identical segments in different orders. While human coders would likely recognize the commonality and judge these as highly similar, the dynamic time warping algorithm is sensitive to ordering and would judge them as very different. So our dynamic time warping analysis may at times judge sequences of gestures highly dissimilar when in fact they are merely ordered differently. There are ways to circumvent this by only looking for trajectory overlaps rather than ordering through time (Pouw & Dixon, 2019)but such analyses goes beyond the current approach. Importantly, this limitation stacks the deck against finding increases in systematicity, so it speaks to the robustness of our measures (and perhaps indicates the limited occurrence of reordering elements) that some degree of systematicity is nonetheless recovered.  
Both of these caveats mean that our approach to kinematics, like all quantitative analyses of human behavior, requires some degree of human oversight (for meaningful implementation) and human insight (for judicious interpretation). When these requirements are met, we believe that our fully reproducible and automatable methods can make important contributions to the systematic study of continuous communicative signals.

## 4.2 Embodied language evolution

In our view the current findings underline an understanding of language evolution as an embodied process. Consider for example our finding that manual utterances simplify, in the way of reducing in size (see also Namboodiripad et al., 2016), in the reduction of submovements, and the decrease in temporal variability of the utterance if it comprised of multiple submovements. This simplification seems to be a reduction in articulatory effort. Making a smoother, smaller, and more rhythmic movements reduces the states that a sensorimotor routine needs to visit (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984). Thus communicative efficiency increased over the generations at the motoric level, an effect not captured by the content-level repetition and gesture length measure used in the original study (Motamedi et al., 2019). Note that kinematic efficiency could also potentially increase learnability and comprehensibility of the gestures. Speech perception in noisy conditions is more optimal when speech is more rhythmic. Speech perception in noisy conditions is more optimal when speech is more rhythmic (Wang, Kong, Zhang, Wu, & Li, 2018). We submit that the optimization of sensorimotor routines of communication is an integral part of the increased efficiency of communication. This is fully in line with early insights on the psychobiology of language by Zipf or Macneilage (**???**; MacNeilage, 2010), which have been overshadowed to some extent by the focus on codes and information content in the wake of Shannon. Our measures show how it is possible to unite fine-grained measures of the biomechanics of evolving continuous signals with information-theoretic notions of like system-level entropy. Low-level biomechanical optimization and code-level systematic optimization should not be treated as categorically distinct processes.  
We find a similar reduction of complexity of pronunciation as in novice learners of ASL. ASL learners spatially reduce their signs as they become more fluent (Lupton & Zelaznik, 1990; Wilbur, 1990). Compound signs also become shorter as multi-component signs merge into efficiently produced single signs, in a way that is mirrored in the data we have analyzed here. As sign systems evolve, suboptimal organizations of submovements give way to more efficient signs, and temporally extended sequences become more coordinated (Bernstein, 1967; Kelso, Tuller, & Harris, 1983). This makes the dynamics of the evolving gesture systems we have studied here similar in some ways to those of full-blown manual language systems, such as ASL.  
Another sign of the maturity of the culturally evolved gestural systems we have studied here is the fact that head movements can function as cues reliably distinguishing paired gestures. This finding resonates with the known grammatic, phonetic, and prosodic functions that head movements have in sign languages such as ASL (Tyrone & Mauk, 2016). Indeed, as Sandler (2018) has argued for sign languages, the expressive power of the body lies in the combination of different articulators, combined into a single synergetic utterance.

## 4.3 Future directions

While we have focused on silent gesture as a test case here, our analyses are applicable to any continuous signaling system. Similar approaches are applied in animal signaling systems, where high dimensional features of tokens are mapped onto lower dimensional space so to identify stably distinct patterns (Sainburg, Thielk, & Gentner, 2019). Staying within the domain of human communication, the current analysis could be applied to speech acoustics, semantics and gesture kinematics in unison (Pouw & Dixon, 2019; Pouw et al., 2021). Indeed, in a study by Perlman, Dale, and Lupyan (2015) it is shown how dynamic aspects of vocalization signaling systems become more efficient, similar to our current reduction in kinematic complexity. These findings, together with work showing the tight connection between speech and gesture (Bosker & Peeters, 2020; Pouw et al., 2020a, 2020b), make it a natural next step to look at multimodal iterated learning experiments. Furthermore, our approach can inform work on communicative alignment in conversation (Rasenberg, Özyürek, & Dingemanse, 2020) or the ways in which people can repeat aspects of each other’s communicative behavior. In short, our analysis here has not only yielded new findings in the cultural evolution of gestural communication systems, but opens the door towards a broader research program in which the biomechanical aspects of communication systems are studied alongside their structural and semiotic aspects.

## 4.4 Conclusion

Human communicative behavior tends to combine categorical elements and continuous properties, but for technological as well as theoretical reasons the categorical elements of evolving linguistic systems have long received more attention than their continuous aspects. Here we have contributed to the study of multimodal language and cognition by considering the gesture kinematics of evolving gestural systems. We have used computer vision techniques to analyze the kinematic properties of evolving gestural systems, showing that over generations of learners, the dynamics of head and upper limb movements become simpler, increase in systematicity, and give rise to kinematic dialects. Our kinematic measures help characterize fine-grained levels of linguistic organization that remain out of reach of content-based discretized coding approaches, providing novel insights that corroborate and complement prior approaches. Our findings providing an unprecedented view of how gestures become structured and increasingly language-like as they evolve, in ways that are directly related to the coordination and simplification of bodily movements. While considerations of communicative efficiency and systematicity have so far been mostly based on analyses of discrete symbol systems like written words and text corpora, our work shows how hallmark features of linguistic systems may be grounded directly in the biomechanical properties of dynamically evolving systems of continuous signals.

# 5 References

Beecks, C., Hassani, M., Brenger, B., Hinnell, J., Schüller, D., Mittelberg, I., & Seidl, T. (2016). Efficient query processing in 3D motion capture gesture databases. *International Journal of Semantic Computing*, *10*(01), 5–25. <https://doi.org/10.1142/S1793351X16400018>

Beecks, C., Hassani, M., Hinnell, J., Schüller, D., Brenger, B., Mittelberg, I., & Seidl, T. (2015). Spatiotemporal similarity search in 3D motion capture gesture streams. In C. Claramunt, M. Schneider, R. C.-W. Wong, L. Xiong, W.-K. Loh, C. Shahabi, & K.-J. Li (Eds.), *Advances in Spatial and Temporal Databases* (pp. 355–372). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-22363-6_19>

Bernstein, N. (1967). *The Co-ordination and Regulations of Movements* ([1st English ed.] edition). Pergamon Press.

Bolinger, D. (1968). *Aspects of Language*. New York: Harcourt, Brace, and World.

Bosker, H. R., & Peeters, D. (2020). Beat gestures influence which speech sounds you hear. *Proceedings of the Royal Society B: Biological Sciences*. <https://doi.org/10.1101/2020.07.13.200543>

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1302–1310). Honolulu, HI: IEEE. <https://doi.org/10.1109/CVPR.2017.143>

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*. <https://doi.org/10.1017/S0140525X1500031X>

Cornish, H., Dale, R., Kirby, S., & Christiansen, M. H. (2017). Sequence Memory Constraints Give Rise to Language-Like Structure through Iterated Learning. *PLoS ONE*, *12*(1), e0168532. <https://doi.org/10.1371/journal.pone.0168532>

Csárdi, G. (2019). Package ’igraph’ Network Analysis and Visualization (Version 1.2.4.1). Retrieved from <http://bioconductor.statistik.tu-dortmund.de/cran/web/packages/igraph/igraph.pdf>

Dale, R., & Kello, C. T. (2018). “How do humans make sense?” Multiscale dynamics and emergent meaning. *New Ideas in Psychology*, *50*, 61–72. <https://doi.org/10.1016/j.newideapsych.2017.09.002>

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>

Donaldson, J. (2016). Tsne: T-Distributed Stochastic Neighbor Embedding for R (t-SNE) (Version 0.1-3). Retrieved from <https://CRAN.R-project.org/package=tsne>

Eagle, N., Macy, M., & Claxton, R. (2010). Network Diversity and Economic Development. *Science*, *328*(5981), 1029–1031. <https://doi.org/10.1126/science.1186605>

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge University Press.

Enfield, N. J. (2014). *Natural causes of language: Frames, biases, and cultural transmission*. Language Science Press. Retrieved from <http://books.google.com?id=RIabBQAAQBAJ>

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture’s form. *Gesture*, *4*(2), 157–195. <https://doi.org/10.1075/gest.4.2.04ger>

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. <https://doi.org/10.1016/j.tics.2019.02.003>

Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in *R* : The **Dtw** Package. *Journal of Statistical Software*, *31*(7). <https://doi.org/10.18637/jss.v031.i07>

Haviland, J. B. (2013). The emerging grammar of nouns in a first generation sign language: Specification, iconicity, and syntax. *Gesture*, *13*(3), 309–353. <https://doi.org/10.1075/gest.13.3.04hav>

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>

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers’ responses. *Journal of Pragmatics*, *43*(14), 3522–3536. <https://doi.org/10.1016/j.pragma.2011.08.002>

Kelso, J. A. S., Tuller, B., & Harris, K. (1983). A “Dynamic Pattern” perspective on the control and coordination of movement. In *The production of speech*. Berlin: Springer-Verlag. Retrieved from <https://link.springer.com/chapter/10.1007/978-1-4613-8202-7_7>

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(6), 812–832.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press. Retrieved from <http://books.google.com?id=hDXnnzmDkOkC>

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. <https://doi.org/10.1016/j.conb.2014.07.014>

Lepic, R., Börstell, C., Belsitzman, G., & Sandler, W. (2016). Taking meaning in hand: Iconic motivations in two-handed signs. *Sign Language & Linguistics*, *19*(1), 37–81. <https://doi.org/10.1075/sll.19.1.02lep>

Lupton, L. K., & Zelaznik, H. N. (1990). Motor Learning in Sign Language Students. *Sign Language Studies*, *1067*(1), 153–174. <https://doi.org/10.1353/sls.1990.0020>

Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605. Retrieved from <http://jmlr.org/papers/v9/vandermaaten08a.html>

MacNeilage, P. F. (2010). *The Origin of Speech*. Oxford University Press. Retrieved from <http://books.google.com?id=BNQUDAAAQBAJ>

McNeill, D. (2005). *Gesture and Thought*. Chicago: The University of Chicago Press. Retrieved from <https://www.press.uchicago.edu/ucp/books/book/chicago/G/bo3633713.html>

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, *192*, 103964. <https://doi.org/10.1016/j.cognition.2019.05.001>

Mueen, A. K., & Keogh, E. (2016). Extracting optimal performance from dynamic time warping. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2129–2130). [https://doi.org/10.1145/ 2939672.2945383](https://doi.org/10.1145/%202939672.2945383)

Muller, M. (2007). *Information retrieval for music and motion*. Heidelberg, Germany: Springer.

Namboodiripad, S., Lenzen, D., Lepic, R., & Verhoef, T. (2016). Measuring conventionalization in the manual modality. *Journal of Language Evolution*, *1*(2), 109–118. <https://doi.org/10.1093/jole/lzw005>

Östling, R., Börstell, C., & Courtaux, S. (2018). Visual Iconicity Across Sign Languages: Large-Scale Automated Video Analysis of Iconic Articulators and Locations. *Frontiers in Psychology*, *9*. <https://doi.org/10.3389/fpsyg.2018.00725>

Padden, C. A., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, *13*(3), 287–308. <https://doi.org/10.1075/gest.13.3.03pad>

Pattee, H. H., & Rączaszek-Leonardi, J. (2012). *LAWS, LANGUAGE and LIFE: Howard Pattee’s classic papers on the physics of symbols with contemporary commentary*. Springer Science & Business Media. Retrieved from <http://books.google.com?id=raEQodcVdYQC>

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, *2*(8), 150152. <https://doi.org/10.1098/rsos.150152>

Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., & Dixon, J. A. (2020a). Gesture-speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*. <https://doi.org/10.1111/nyas.14532>

Pouw, W., & Dixon, J. A. (2019). Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles. *Discourse Processes*, *57*(4), 301–319. <https://doi.org/10.1080/0163853X.2019.1678967>

Pouw, W., Harrison, S. J., Esteve-Gibert, N., & Dixon, J. A. (2020b). Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. *The Journal of the Acoustical Society of America*, *148*(3), 1231–1247. <https://doi.org/10.1121/10.0001730>

Pouw, W., & Trujillo, J. P. (2019). *Materials Tutorial Gespin2019 - Using video-based motion tracking to quantify speech-gesture synchrony*. Retrieved from <10.17605/OSF.IO/RXB8J>

Pouw, W., Wit, J., Bögels, S., Rasenberg, M., Milivojevic, B., & Ozyurek, A. (2021). Semantically related gestures move alike: Towards a distributional semantics of gesture kinematics. *Proceedings of the 23rd International Conference on Human-Computer Interaction*. <https://doi.org/10.31219/osf.io/pgq6m>

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*(11), e12911. <https://doi.org/10.1111/cogs.12911>

Ravignani, A., Delgado, T., & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, *1*(1, 1), 1–7. <https://doi.org/10.1038/s41562-016-0007>

Rączaszek-Leonardi, J., & Kelso, S. J. A. (2008). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology*, *26*(2), 193–207. <https://doi.org/10.1016/j.newideapsych.2007.07.003>

Ripperda, J., Drijvers, L., & Holler, J. (2020). Speeding up the detection of non-iconic and iconic gestures (SPUDNIG): A toolkit for the automatic detection of hand movements and gestures in video data. *Behavior Research Methods*, *52*(4), 1783–1794. <https://doi.org/10.3758/s13428-020-01350-2>

Sainburg, T., Thielk, M., & Gentner, T. Q. (2019). Latent space visualization, characterization, and generation of diverse vocal communication signals. *bioRxiv*, 870311. <https://doi.org/10.1101/870311>

Sandler, W. (2018). The Body as Evidence for the Nature of Language. *Frontiers in Psychology*, *9*. <https://doi.org/10.3389/fpsyg.2018.01782>

Sato, A., Schouwstra, M., Flaherty, M., & Kirby, S. (2020). Do all aspects of learning benefit from iconicity? Evidence from motion capture. *Language and Cognition*, *12*(1), 36–55. <https://doi.org/10.1017/langcog.2019.37>

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in nicaragua. *Science*, *305*(5691), 1779–1782. <https://doi.org/10.1126/science.1100199>

Trettenbrein, P. C., & Zaccarella, E. (2021). Controlling Video Stimuli in Sign Language and Gesture Research: The OpenPoseR Package for Analyzing OpenPose Motion-Tracking Data in R. *Frontiers in Psychology*, *12*. <https://doi.org/10.3389/fpsyg.2021.628728>

Trujillo, J. P., Vaitonyte, J., Simanova, I., & Özyürek, A. (2019). Toward the markerless and automatic analysis of kinematic features: A toolkit for gesture and movement research. *Behavior Research Methods*, *51*(2), 769–777. <https://doi.org/10.3758/s13428-018-1086-8>

Tyrone, M. E., & Mauk, C. E. (2016). The Phonetics of Head and Body Movement in the Realization of American Sign Language Signs. *Phonetica*, *73*(2), 120–140. <https://doi.org/10.1159/000443836>

Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57–68. <https://doi.org/10.1016/j.wocn.2014.02.005>

Verhoef, T., Kirby, S., & de Boer, B. (2016). Iconicity and the Emergence of Combinatorial Structure in Language. *Cognitive Science*, *40*(8), 1969–1994. <https://doi.org/10.1111/cogs.12326>

Wang, M., Kong, L., Zhang, C., Wu, X., & Li, L. (2018). Speaking rhythmically improves speech recognition under "cocktail-party" conditions. *The Journal of the Acoustical Society of America*, *143*(4), EL255. <https://doi.org/10.1121/1.5030518>

Wilbur, R. B. (1990). An experimental investigation of stressed sign production. *International Journal of Sign Linguistics*, *1*(1).

Yadav, A. K., Tomar, D., & Agarwal, S. (2013). Clustering of lung cancer data using Foggy K-means. In *2013 International Conference on Recent Trends in Information Technology (ICRTIT)* (pp. 13–18). <https://doi.org/10.1109/ICRTIT.2013.6844173>

# 6 Supplemental

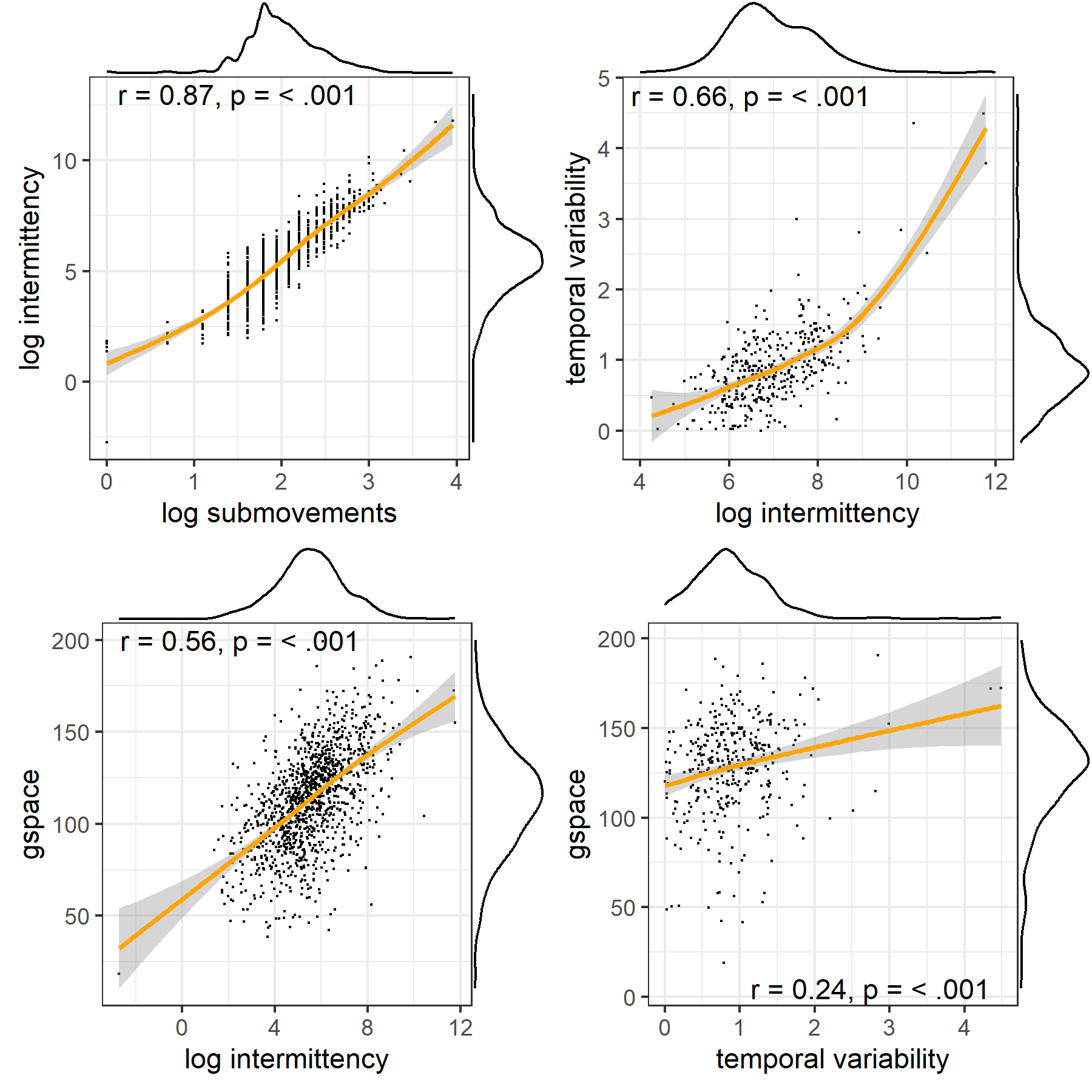
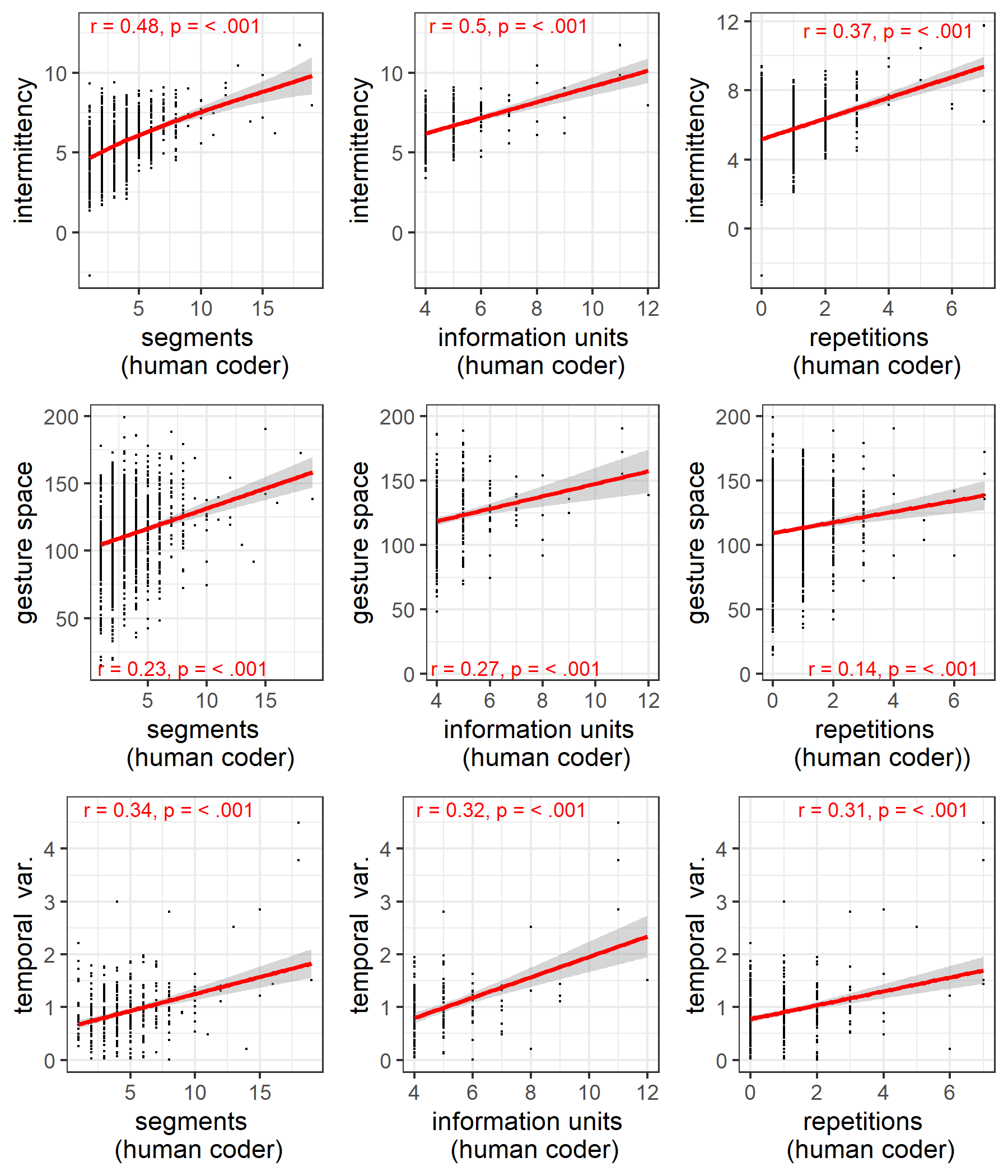
Figure S1. Correlations and distributions for kinematic measures per trial  *Note Figure S1*. Left upper panel, correlations and distributions are shown for intermittency and submovement. Given their high correlation we will use intermittency score for our final analysis. Other correlations are shown for the selected measures, rhythmiticy, gesture space and intermittency.

Figure S2. Correlations of kinematic measures with human-coded gesture information  
 *Note figure S2.* On the horizontal axes the human-coded number of gesture segments, unique information units, and the number repetitions (of information units) are shown. On the vertical axes our automatic kinematic measures are shown: intermittency, gesture space, and temporal variability. The findings show that our measures are a proxy for human judgments, such that more intermittent kinematics reflect more information units (repeater and/or unique). Larger gesture space is related to more information units. Lower temporal variability in kinematics is further associated with more information units.