Multiscale kinematic analysis reveals structural properties of change 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 Language and Cognition Lab, University College London

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

In the current approach we perform a multiscale quantification of kinematic changes of an evolving silent gesture system. Silent gestures consist of complex multi-articulatory movement that have so far proven elusive to quantify in a structural and reproducable way, and is primarily studied through human coders meticulously interpreting the referential content of gestures. Here we demonstrate that the signal’s objective form (kinematics of gesture) can however also be investigated for the emergence of linguistic constraints. Here we reanalyzed the video data from a gesture evolution experiment (Motamedi et al. 2019), which originally showed increases in systematicity in gesture’s content over time. We applied a signal-based approach, utilizing computer vision techniques to quantify kinematics. Then we performed a kinematic analysis, showing that over generations of language users, silent gestures became more efficient and less complex in their kinematics. We further detect systematicity of the gesture form on the level of their interrelations, which proved itself as a proxy of systematicity obtained via human observation data, and was related to the kinematic changes observed on the token level. Finally we demonstrate that gesture kinematic cultures emerge over generations, as 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 gesture forms, paving a way towards a reproducible analysis of linguistic constraints in continuous multidimensional signals.

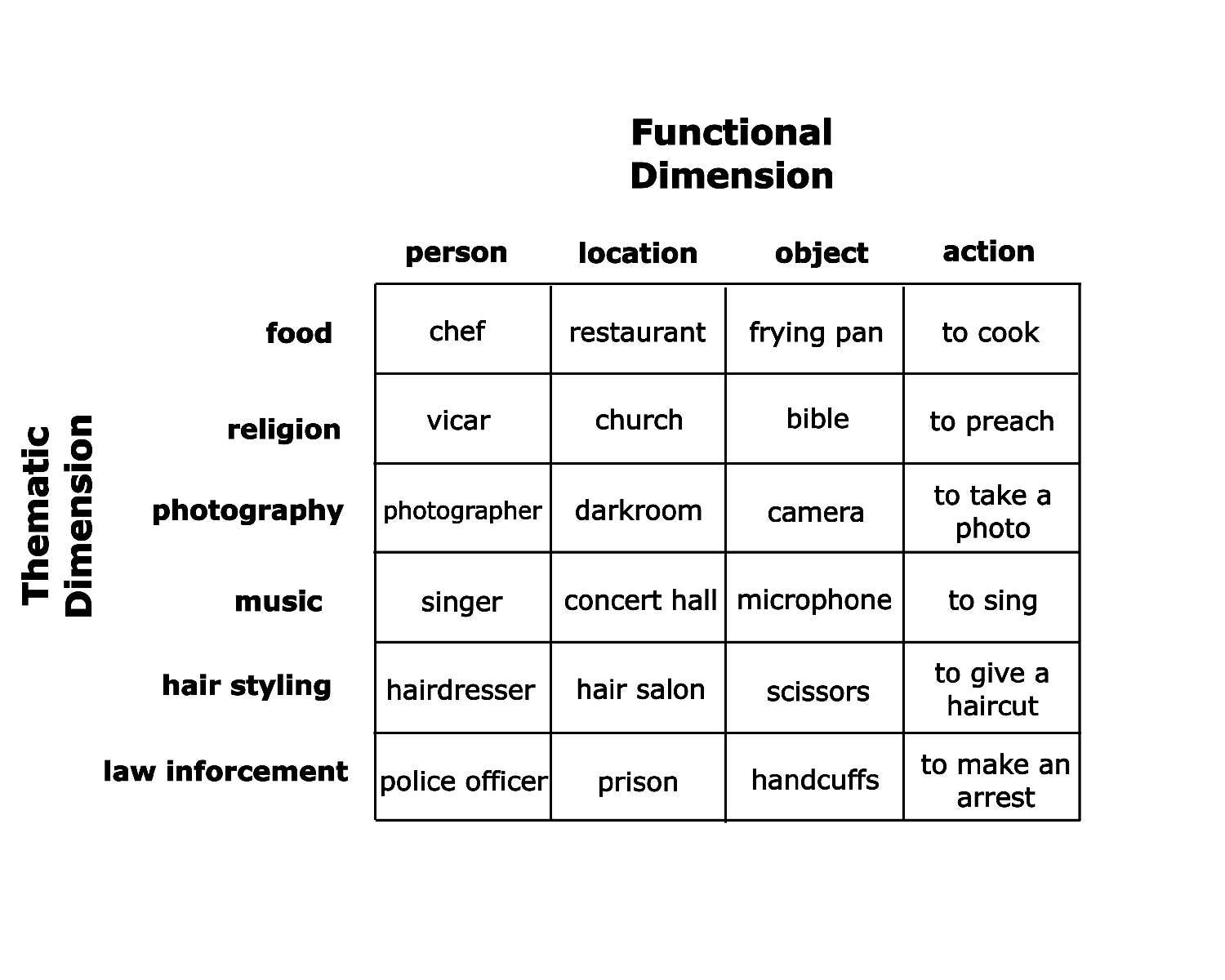
*Keywords:* language evolution, silent gesture, kinematics, systematicity

*Word count:* X

Multiscale kinematic analysis reveals structural properties of change in evolving manual languages in the lab

How do language scientists judge a communicative signal to be linguistic in nature? What properties must a communicative signal have? It can be contended that no *particular* property in a communicative signal is revealing about its systematic function. This is where the notion of arbitrariness comes from: when there is an arbitrary mapping between a signal’s form and its meaning, we cannot attune to the meaning of the signal in reference to its form alone. We can only attune to such a meaning when knowledgeable about how a signal’s form functions in relation to other signaling tendencies in relevant contexts.  
Arbitrariness in our view is a relational concept and has many forms. Most communicative signals have forms that *seem* arbitrary to their meaning if we are ignorant about how they relate to physical aspects of the signaling modality (e.g., emotional stress and vocal fold tensioning), or when we are ignorant about how they relate to common action routines and event experiences as in the case of iconic reference. Similarly, some signals seem arbitrary when not properly related to linguistic constraints (e.g., syntactic rules of word class order). Importantly, arbitrariness is not something that is fixed, it can be *resolved* by relating a signals form to relevant aspects of the systems that constrain them. A signals form can be constrained by systems that operate on different timescales. Vocal fold tensioning due to emotional stress can be said to operate more directly on signaling, but we cannot say the same for linguistic constraints. In the linguistic case, the form of a communicative signal is constrained by a culturally shaped system operating on a timescale that transcends the level at which those constraints are exerted (Pattee & Rączaszek-Leonardi, 2012; Rączaszek-Leonardi & Scott Kelso, 2008).  
Importantly, in this view, arbitrariness only means we have not properly related a signal’s form with the systems they are constrained by, and the timescales such systems operate on. Arbitrariness can lead us astray in this way by leading to the conclusion that a signals form *is* arbitrary and not revealing of linguistic competences. A paradigmatic example of this is the current mainstream to study manual gesture communication. Manual gestures are seldomly semiotically studied based on the signal’s measurable form, namely manual and whole-body postures in movement (i.e., the kinematic level). Instead gestures are studied as already categorizable expressions by researchers inferring meaning from their form. As such the kinematics are in one sense reduced (from continuous to discrete tokens) and another sense enriched (from movement to message) with meanings that are *projected by human coders*. This is indeed the only viable way if we agree that kinematics *are* arbitrary, as the very movement as such is not what is important, but what the movement signifies. This seems supported by the very fact that a gesture may mean something completely different while being similar in kinematics.  
In the current approach we will assess whether there is a way to study communicative body movements revealing of the linguistic system they are possibly constrained by. Communicative body movements, or silent gestures, are an exemplary test case for the current view that linguistic constraints are relationally present in the signal’s form. This is because, as mentioned, a semiotic study of gesture kinematics is underdeveloped (but see Lepic, Börstell, Belsitzman, & Sandler, 2016; Östling, Börstell, & Courtaux, 2018), and relatedly, because gesture kinematics are highly complex signals, involving multiple body parts and continuous trajectories that are generally held to be difficult to parse into discrete meanings (see e.g., Sandler, 2018).

#Context of the present study  
There is an interdisciplinary scientific effort to unveil the historical and/or necessary constraints that allow(ed) for the emergence of human language (e.g., Bickerton, 2009; Deutscher, 2005; Fitch, 2010, 2010; McNeilage, 2008; Tomasello, 2008). An important approach within this enterprise comes from experimental cognitive science (Scott-phillips & Kirby, 2010), where interactive communication processes likely to have contributed to language emergence are simulated in the lab (e.g., Kirby, Cornish, & Smith, 2008).  
In this iterated learning approach to language evolution, agents are tasked to learn a novel set of signals, iteratively transmitted to later generations and/or used in communication by later generations (iterated learning + communication). Over many cycles of learning and use, the signals are affected by various transmission biases (e.g., Christiansen & Chater, 2016; Enfield, 2016). Processes of iterated learning and communication can simulate how structural properties such as systematicitity, learnability, and compositionality evolve from more simpler communication systems — a process that must have occurred in human language evolution too (Bickerton, 2009). 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 is able to tweak itself 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 limited 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 signals (Cornish, Dale, Kirby, & Christiansen, 2017; Ravignani, Delgado, & Kirby, 2016; Verhoef, Kirby, & de Boer, 2016).  
As mentioned communicative body movements are challenging in this way, but insights from meticulous study by (psycho-)linguists have revealed important insights on how they evolve. For example, in dyadic social interaction, repeated gestural referrals to an object or a picture will lead to those gestures becoming more reduced in size (Gerwing & Bavelas, 2004; Namboodiripad, Lenzen, Lepic, & Verhoef, 2016). This is comparable to research in ‘pictionary’ paradigms where a concept is drawn out and to be interpreted by another player. After repeated trials of drawing, a reduction of the drawings’ complexity is observed, with smaller-sized and less iconic drawings as a result, while communicative accuracy increases over time (Fay, Garrod, Roberts, & Swoboda, 2010; Garrod, Fay, Lee, Oberlander, & Macleod, 2007). Though, drifts from less or more iconicity are not fixed processes. When interaction between people is opened up, a whole new suit of social affordances arise. For example, while gestures may reduce in size and iconicity when some common ground is established, at any moment an interlocutor may request a clarification, soliciting large and iconic gestures per implicitly requested (Bavelas, Gerwing, Sutton, & Prevost, 2008; Holler & Wilkin, 2011). In such moments of interactional repair, common ground is calibrated and re-established. Indeed, establishing common ground is not something that is a linear phenomenon, but an interactive and dialectic process. These and many other interactive affordances turn out to be of central importance for smooth everyday language use (Dingemanse, Roberts, et al., 2015), and according to cultural evolutionary accounts of language, such local-scale processes of interaction and transmission between communicators are crucial for the emergence of any linguistic system (e.g., Enfield, 2016; Kirby & Christiansen, 2003; Kirby, Griffiths, & Smith, 2014; Raviv, Meyer, & Lev-Ari, 2019). A key question that drives cultural evolution research is which particular interactive constraints produce pressures for a certain communication system to adapt in one way or another, and how effective solutions are negotiated at the possible expense of other communicatively efficient solutions (Dingemanse et al., 2015).  
In a recent iterated learning study with silent gesture (Motamedi, Schouwstra, Smith, Culbertson, & Kirby, 2019) two such possible constraints, transmission across generations and communication within generations, were studied simultaneously as well as separately. Learning occurred with a set of silent gesture-concept mappings (i.e., communicative tokens) communicated through five iterations of vertical transmissions, where gestures were transmitted from one participant to-be-reproduced by the next participant. Or, tokens would be communicated through five horizontal interactions in a director-matcher type task. These constraints - interaction and transmission - were first studied in combination in experiment 1, which is the focus for the current paper. An important aspect of the study was that every concept was characterisable along two 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  


These dimensions provided possible axes for compressibility of the communicative tokens. After all, by combining 10 unique gestures one can pick out any referent (e.g., “to make an arrest”) from the 24 token meaning space, one gesture marking the functional category (e.g., “action”) and another gesture for the theme category (“justice”). To exemplify further, once confronted with communicating 24 meanings one can invent 24 unique gesture utterances, such as in the following videos for “to sing” (<https://osf.io/d8srx/>) and “singer” (<https://osf.io/974ke/>), which is challenging to do since they are both very much related. However, one can also 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/>”), and then followed by the same thematic marking gesture conveying “music”. Not only do these general functional markers aid the disambiguation of related meanings, this expressive invention also allows for a systematic reemployment of functional markings for the whole meaning space through compositionality. Once you invent 4 functional marker gestures, and 6 thematic marker gestures, you can systematically recombine these to convey 24 meanings. The communicative system then has compressed its information density from 24 information units to 10 information units.  
Motamedi and colleagues (2019) indeed qualitatively observed such signs of compression of the meaning space as the system developed. In early iterations of learning, large-sized iconic enactments were the most common way of gesturally depicting the referents. However, the occurrences of functional markers increased over generations, which represented meaning components reused across gestures. This kind of functional marking mainly targeted the thematic and function categories.  
With meticulous hand coding of the different referential components of each silent gesture, it could further be quantitatively tested whether there was indeed systematicity emerging. The gesture coding included information about 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, Shannon entropy was computed, which expresses compressibility of the content of a signal, i.e., the amount of information that is needed to compress the signal. When a lot of referential components in the gesture utterance recur between other gestures that the participant produced, the gesture system has a more compressible structure and indicates systematic reuse of gestural components (e.g., Gibson et al., 2019). Dovetailing with the qualitative observations and other quantitative studies in this field (e.g., Verhoef et al., 2016), it was found that gesture-component entropy decreased over the generations. Furthermore, the gestures were coded for the amount of marking for the functional category, and this showed that such gestures 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 predictions that more mature communication systems tend towards maximal efficiency (Gibson et al., 2019), and that gestures tend to simplify in form when repeatedly used (e.g., Gerwing & Bavelas, 2004).  
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 affords novel combinatoriality and therefore increases generativity of a language. It expresses the meaning space with fewer means similar to how 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”).

# 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 et al., 2016; Ripperda, Drijvers, & Holler, 2020). We then performed ‘gesture network analysis’ (Pouw & Dixon, 2019), which is a procedure that combines bivariate time series analysis (Dynamic Time Warping) with network analysis and visualization. Through this gesture network analysis we show that the study of gesture’s form can be revealing of linguistic constraints if we study the kinematic system as a whole, as proposed by the gesture network analysis approach.

We hypothesized the following. Over the iterations:

1. gesture kinematics simplifies (token level).
2. gesture kinematics relationships become more systematic, and this scales with (system level).
3. changes in gesture kinematics are related to the systematicity of gesture (token-system level).
4. idiosyncratic gesture cultures emerge as evidenced by drifts away from the community as a whole (inter-system level).

Importantly, 1) is based on previous research showing how gestures simplify over repeated use (Gerwing & Bavelas, 2004; Namboodiripad et al., 2016), though at present we know of no quantitative kinematic analysis of the simplification of gesture systems as will be performed here. We will use three kinematic properties that revealing of simplication, gesture size, gesture’s intermittency, and gesture rhythm. Whatever kinematic changes we observe, they are not in and of itself evidence for linguistic constraints. Only, 2) and 3) can be revealing of such constraints. Specifically, we will assess whether a Shannon-based Entropy measure computed on kinematic relationships show that the systems becomes more structured (i.e., compressible). By using a conceptually similar measure as has been applied on human coding (Motamedi et al., 2019), we can then also compare whether kinematics’ entropy is related to the content’s entropy. If so, we have a good evidence that linguistic constraints can be objectively studied from systematic changes in gesture form. If we further find that kinematic changes are related to order emerging on the system level, we show how the simplification of gesture kinematics is a co-constitutive (or an embodiment) of order on the system level. Finally, with 4), we show the true potential of a big data implementation to the gesture network approach, by assessing whether as a communicative gesture evolves within a chain of users, that community of users will drift away from other chains of users. In this way we can show how dialects emerge out of the unique historicity built within chains, leading chains to diverge on the whole. Order on the kinematic system level then, is order of an idiosyncratic kind.

# 2 Method

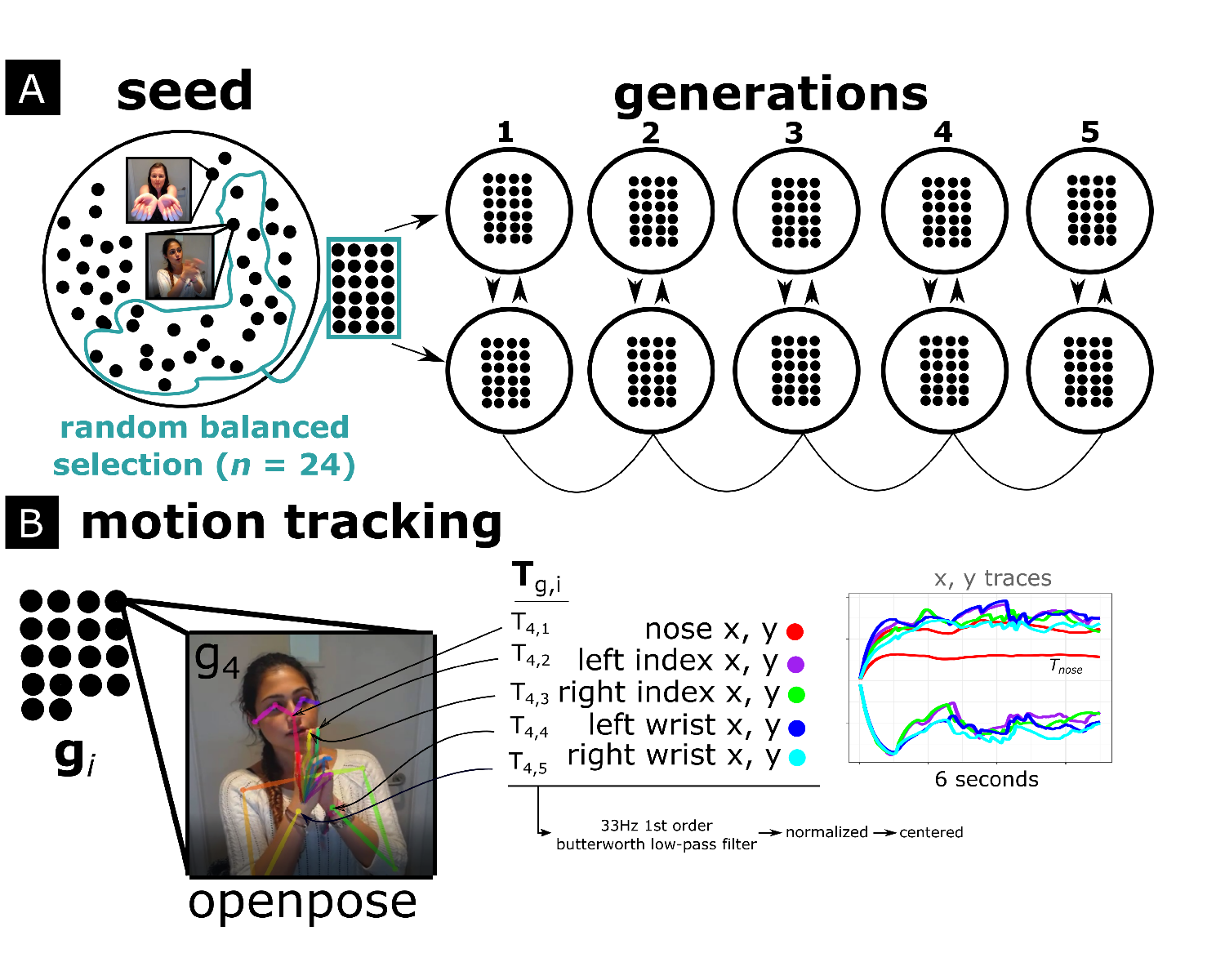
We will follow a bottom-up approach to the study of kinematics as communicative systems, by first showing what specific changes occur in the kinematics of the gestures. For the step 2, we asses possible systematic interrelationships in kinematic patterns of gestures through gesture network analysis (Pouw & Dixon, 2019). Figure 2 and 3 shows the general overview of the pre-processing steps for this experiment. We will discuss each step in this procedure in the following sections, and finally discuss our main gesture network measure (entropy). In the the supplemental information we reserve extra space for sanity checks and graphical descriptive results of the key measures.

### 2.0.1 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 spaces, followed by feedback about their performance. They were then asked to self-record their own copy 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 (withou 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.

### 2.0.2 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, wrists, as well as the nose. For all position traces and its derivatives, we applied 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.

Figure 2. Design experiment and openpose tracking  Note Figure 2\*. The general procedure is shown for the current gesture network analysis. 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; subsequently, 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 () of the nose, the wrists and index fingers. After motion tracking the next steps were dynamic time warping and gesture network analysis (Figure 3).

## 2.1 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 keypoint’ time series seperately 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 gesture segmentation (intermittency score), gesture salience (gesture space), gesture’s temporality (temporal variability). Correlations and distributions are shown in Figure 7.

### 2.1.1 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 keypoint multiplied by the maximum horizontal amplitude, i.e., the area in pixels that has been maximally covered by the movement.

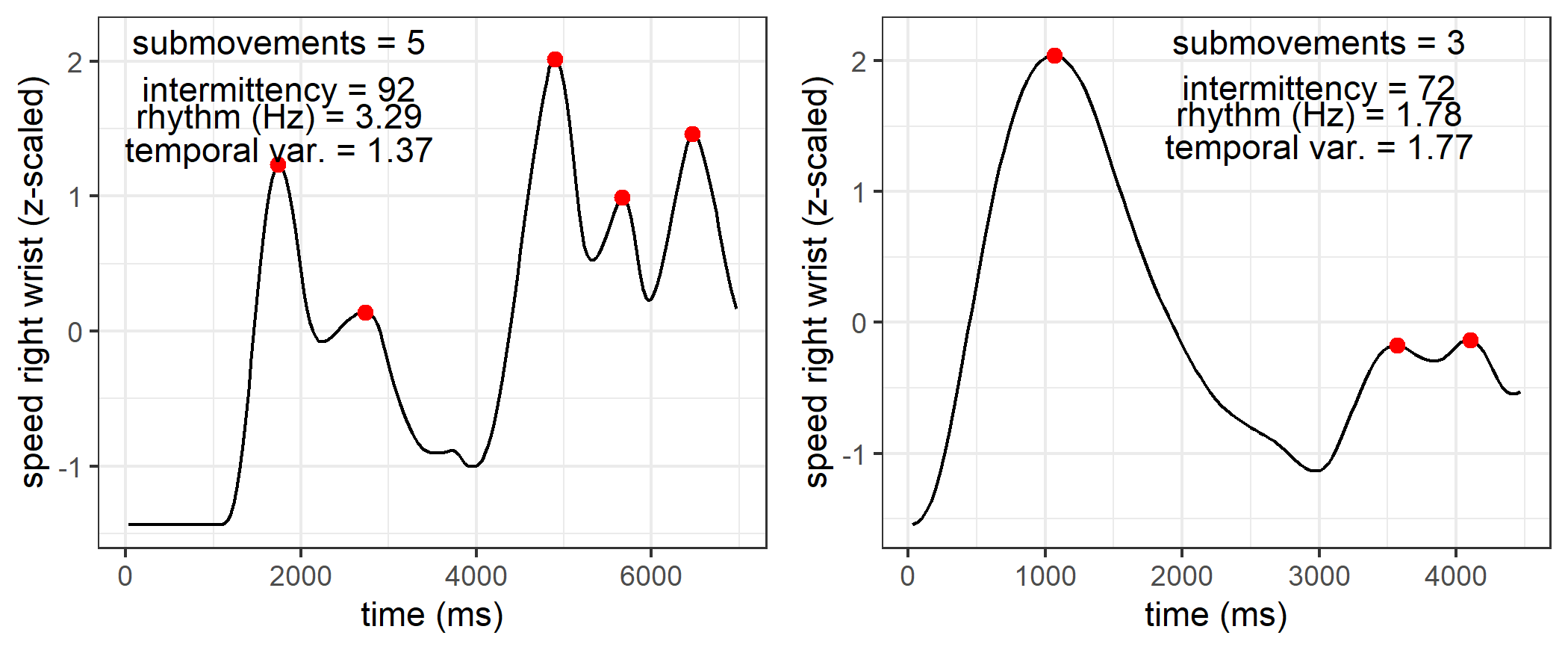
### 2.1.2 Gesture segmentation.

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 interpeak distance at 8 frames, and minimum height = -1 (z-scaled; 1 std.), minimum rise = 0.1 (z-scaled).  
A 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 . As supplemental figure S2 shows, 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 logtransformed our smoothness measures due to skewed distributions.

### 2.1.3 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, i.e., the temporal variability (the opposite of isochrony) 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, this measure cannot be calculated when there are less than 3 submovements (i.e., when there no intervals to detect the temporal variability of).

Figure 3. Overview kinematic measures  *Note Figure 3*. Two time series (belonging to two unique trials) are shown for right-hand wrist speed. From these time series, as well as the time series for other body parts, we computed measures tracking segmentation, namely, submovements (number of observed peaks in red) and intermittency. We further computed measures concerning temporality, namely the average time between submovements, i.e., rhythm in Hertz. We also computed temporal variability, which is the standard deviation of the rhythm in Hertz. Gesture space was calculated from the x,y position traces and is not shown here.

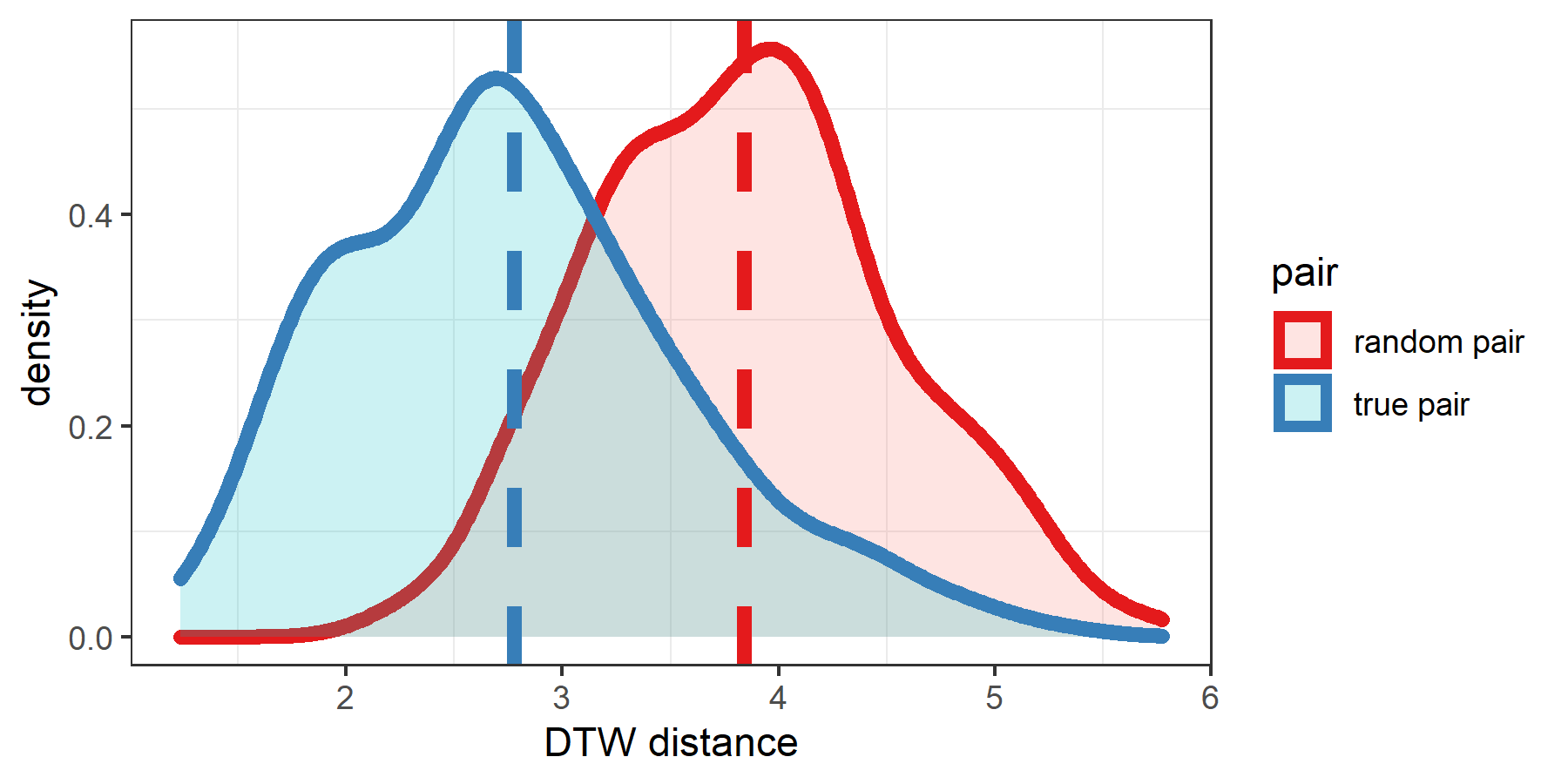
#### 2.1.3.1 Human coding and kinematic measures.

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 are reliably correlating with kinematic intermittency (more intermittent more information), gesture space (larger space more information) and temporal variability (more stable rhythm more segments).

### 2.1.4 Dynamic Time Warping (DTW).

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 need to be stretched and warped 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 if required, yielding a multivariate dependent variant of DTW. We opt for a dependent DTW procedure here as x an 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 keypoint. To compute a single distance measure between gestures, we computed for each gesture comparison a multivariate dependent DTW Distance measure per keypoint, which was then summed for all keypoint comparisons to obtain a single Distance measure D (illustrated in Figure 2C). The D measure thus reflects a general dissimilarity (higher D) or similarity (lower D) of the whole manual+head movement utterance versus another utterance.  
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 & Dixon, 2019).

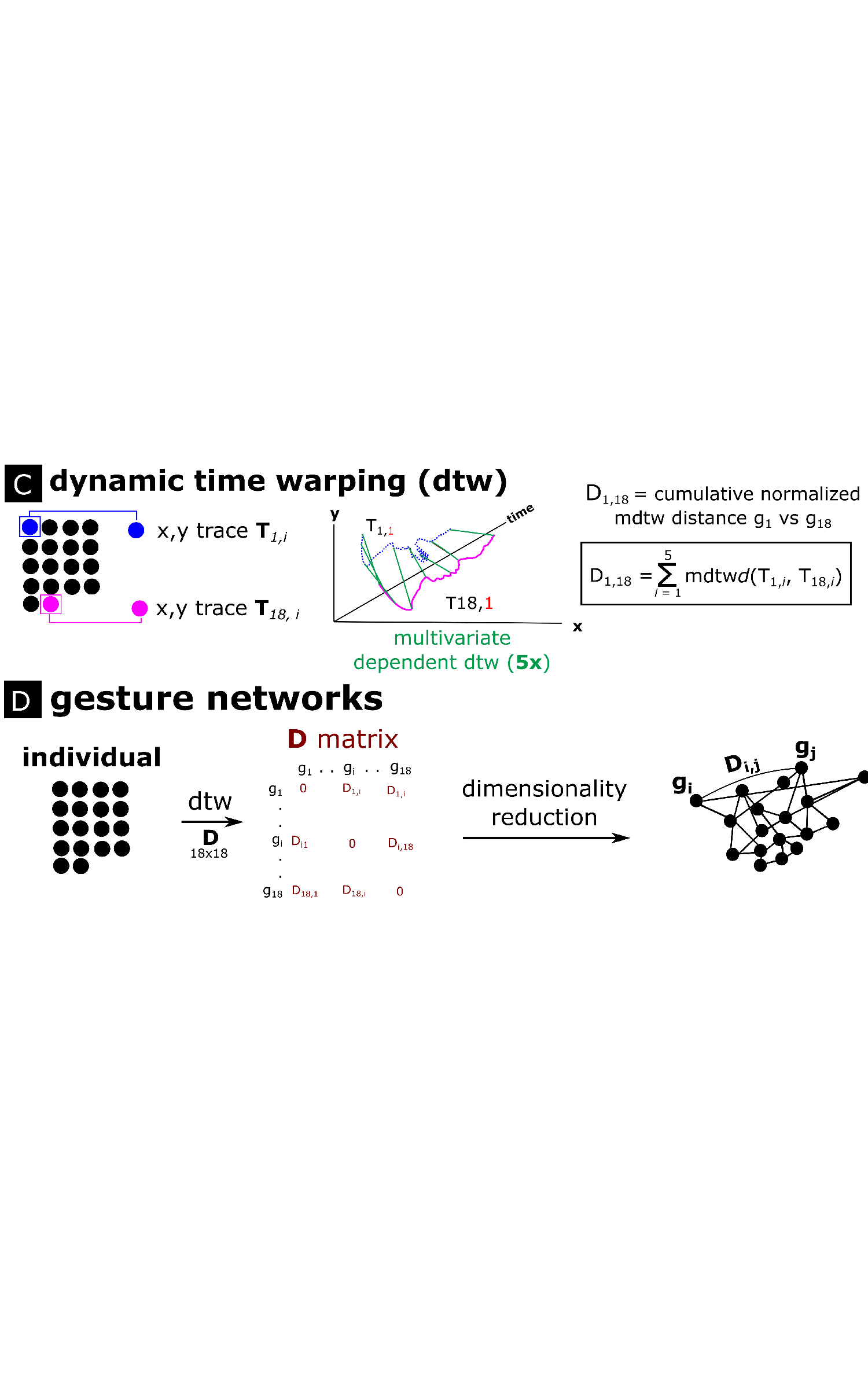
As a demonstration that our D measure reflects actual differences in kinematics, we computed for each individual in each chain the difference between a gesture seed and the gesture that the individual produced to copy it, for generation 1. These “true pairs” must be maximally similar (lower D) as the individual produced their copied gesture short 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).  
Importantly, we also 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 an interesting 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.

### 2.1.5 Gesture networks.

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 distance matrix **D**. 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 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 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).

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 multidimensional scaling, such that nodes indicate gesture utterances and the distance (or weight) between gesture nodes representing the ‘D’ measure, indicating dissimilarity.

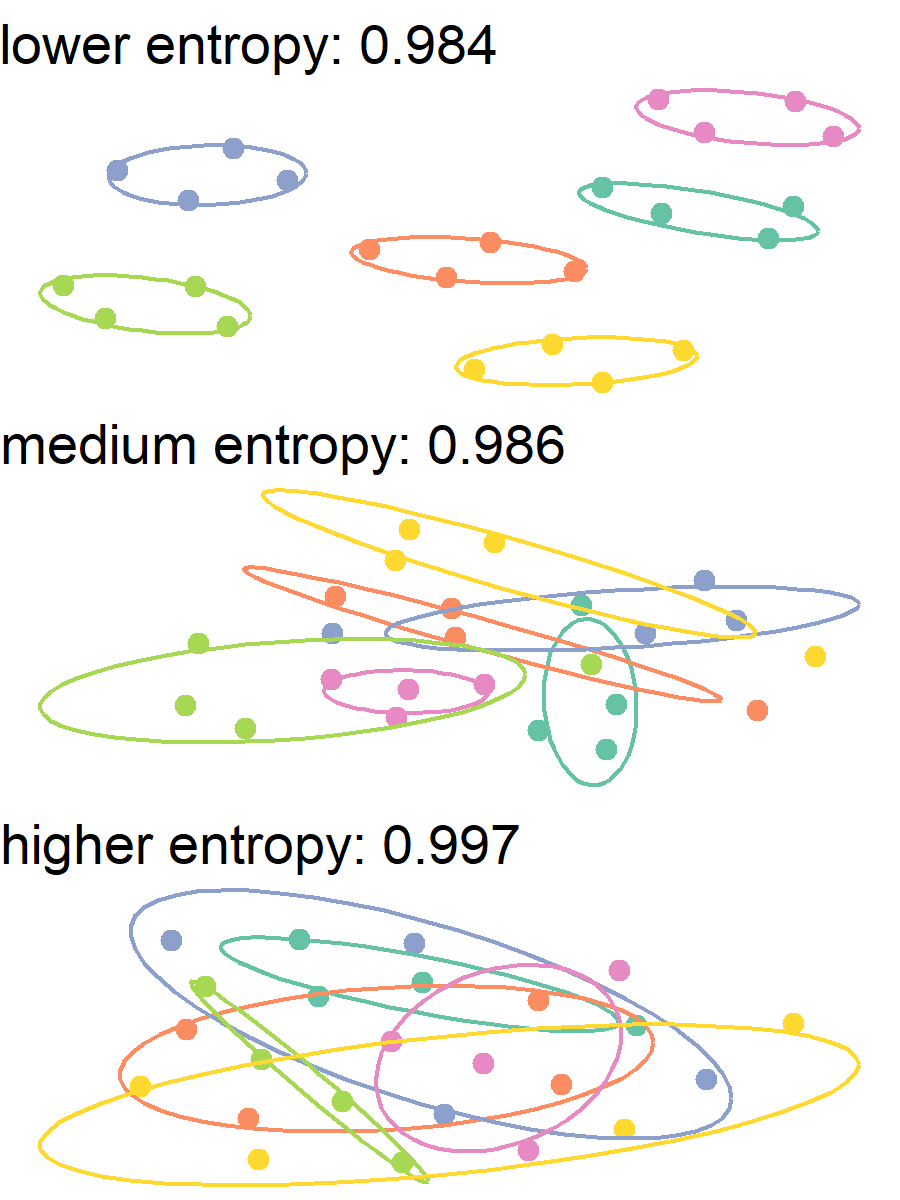
## 2.2 Gesture Network Properties

### 2.2.1 Kinematic Network 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’). If they are more similar, this means lower entropy reflecting that communicative tokens relate in more structural ways to each other. Thus it is important to emphasize here that network entropy gauges in our case how the kinematics interrelationships are compressible (more predictable), and this is conceptually similar as gauging 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 weights of the networks’ edges for each node relative to the shortest path to the other nodes (i.e., connections), and then normalized by the number of connections. In our case the measure quantifies the topological diversity of the gesture relationships, where a lower score indicates more similar relationships and a higher score indicates a more random set of relationships. Specifically, for each gesture node 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 proportial distance:

Here is the distance between gesture and gesture and the total distance involving gesture .  
Figure 4 shows a graphical example.

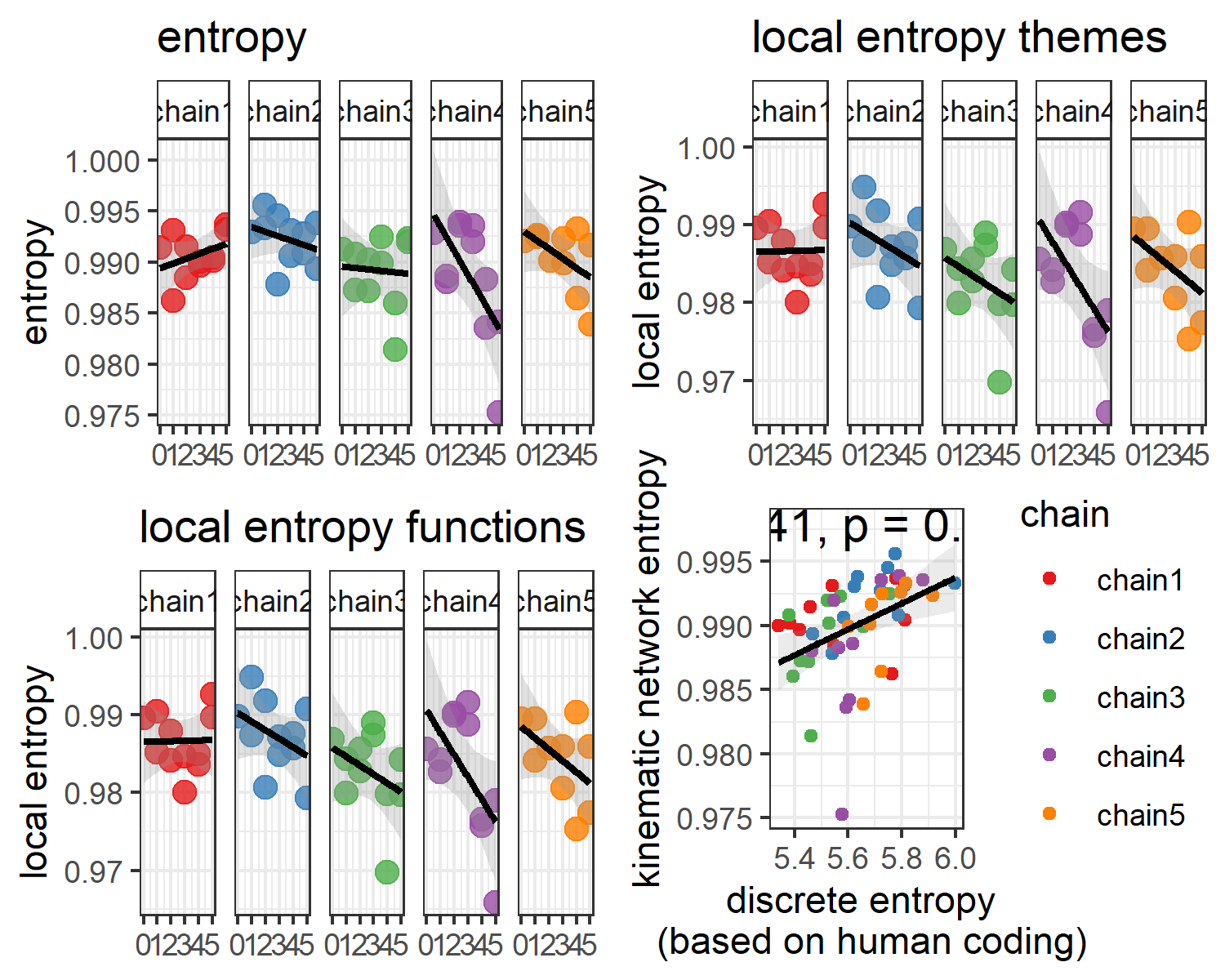
Figure 6. Example network entropy  *Note*.

# 3 Main Results

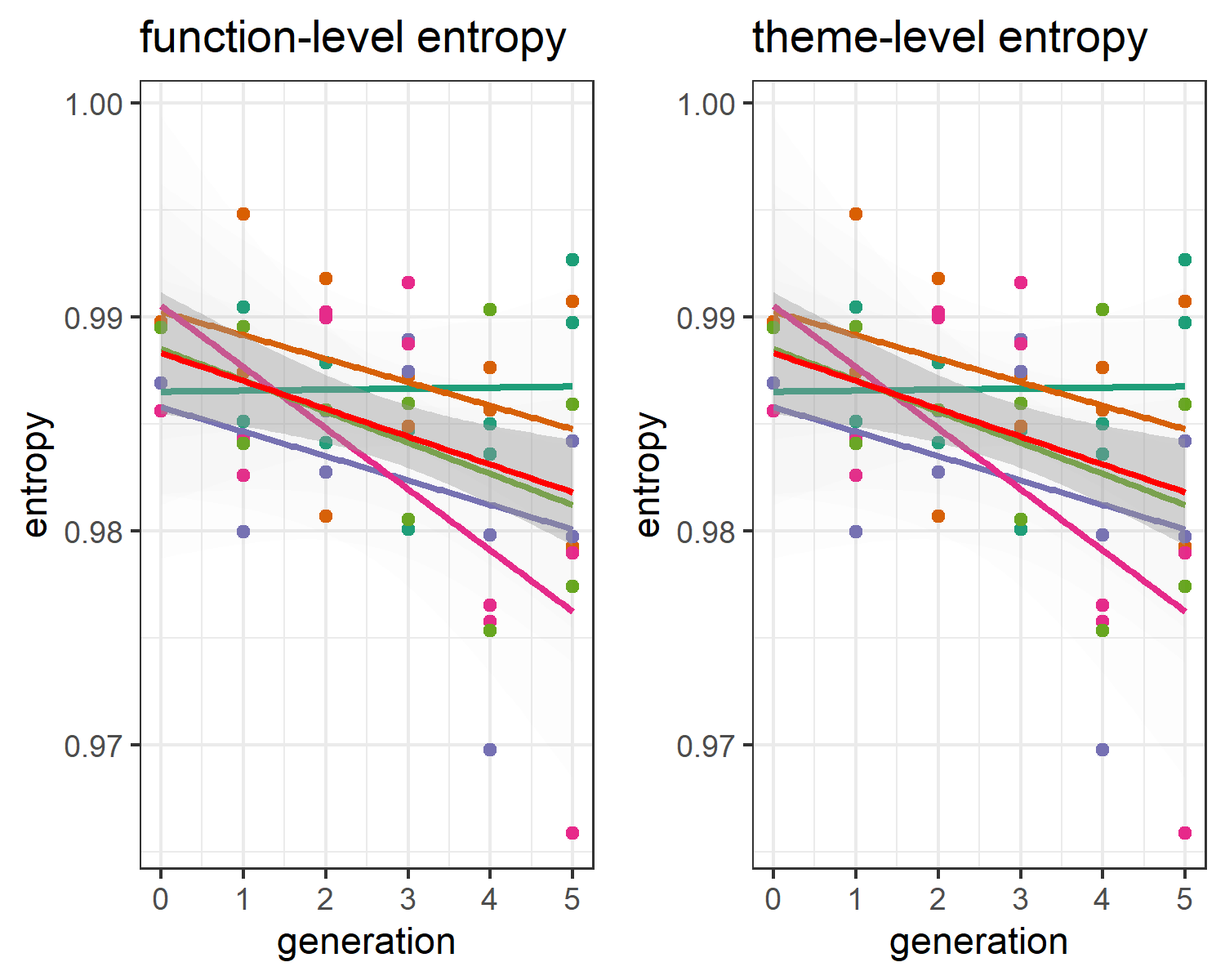
We will first report findings on how relations between communicative tokens changed over the generations, as indicated by our network measures. We then validate whether network entropy approximates systematicity as observed by human coders. Subsequently, we will assess whether network changes occurred between particular tokens, namely the function vs. theme grouping. Finally, we will report on whether structural kinematic changes occurred over the generations for verb (action) and non-verb gestures (objects, persons, locations), and how such kinematic changes related to changes on the network level.

## 3.1 Network changes over generations

Figure 9 shows that for the gesture networks, that entropy was generally decreasing as a function of generation, indicating lower complexity of gesture interrelations as the system matures. Furthermore, there was less clustering at later generations (lower cluster persistence), indicating that kinematic patterns became more differientiable.

Figure 9. Changes in networks measures over generations within chains  *Note Figure 9.* For each chain the changes over generations in entropy and cluster persistence is shown, 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. Two example data points (red, and blue) are shown with their corresponding red and blue network representation (lower panel). In general cluster persistence decreased, indicating less differentiability between tokens. This may be seen in example A where there are relatively large cavities between tokens, while in example B the token organization is more homegonously tessellated. Indeed, entropy tends to decline over the generations, indicating that relationships between tokens became less diverse, possibly indicating systematicity in the way nodes are connected. The right lower panel shows that there there is a strong relationship between the gesture network entropy with that of entropy computed on human-coded information units. That network entropy is uniquely related to systematicity is further corroborated by the finding that cluster persistence is not reliably correlated with entropy based on human coding. It does seem that gesture network analysis is a form-based proxy for systematicity in silent gesture.

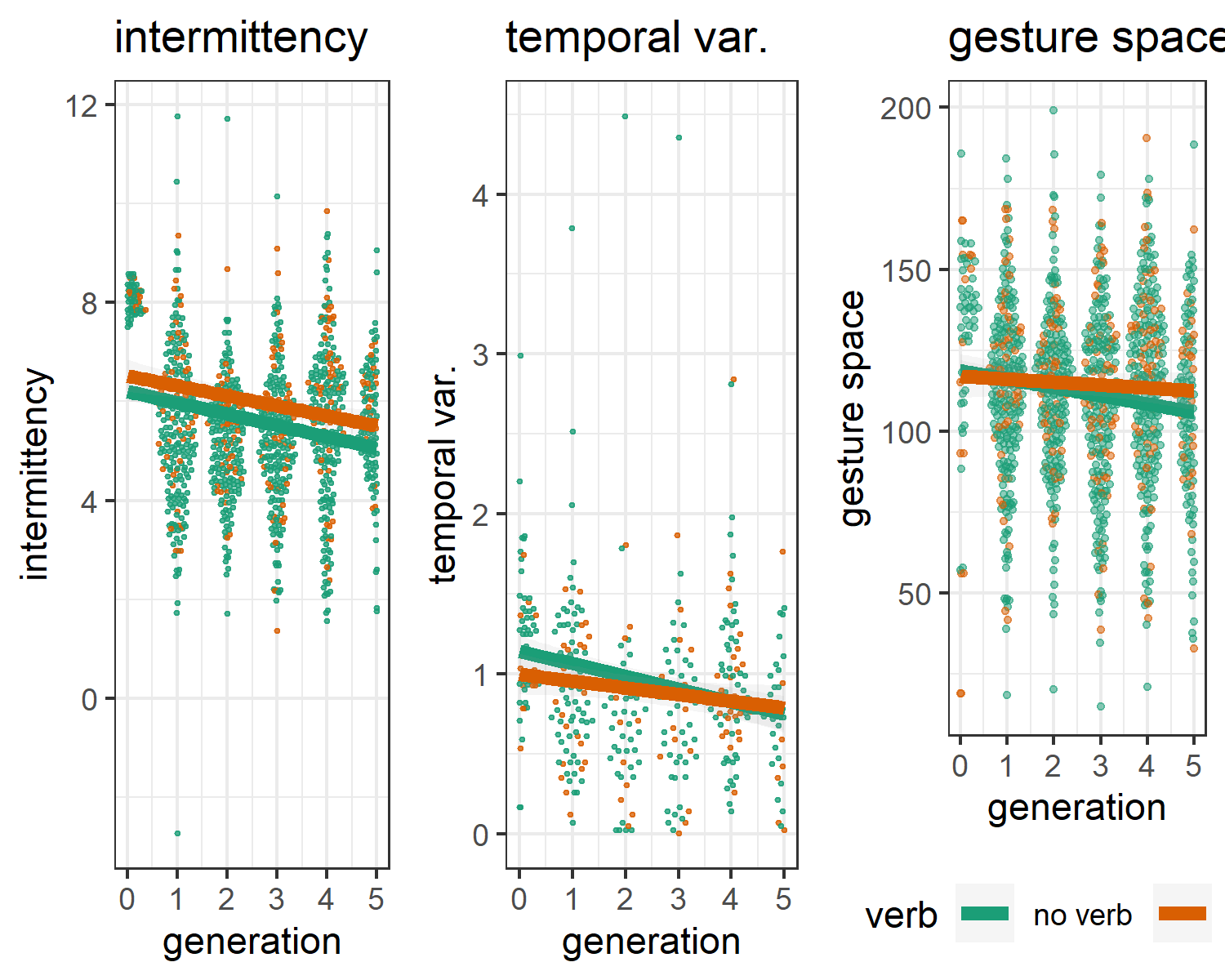
We tested these trends separately for each network property with mixed linear regression models, with chain as random intercept (random slopes did not converge for these models) and generation as independent predictor (0-5 generations, with generation 0 being the seed gesture network).  
Generation was a reliable predictor for entropy as compared to a basemodel predicting the overall mean, chi-squared change (1) = 4.75, *p* = 0.03, model *R*-squared = 0.08. Model estimates showed that with increased generation the entropy decreased, *b* estimate = -0.0006, *t* ( 48.00 ) = -2.19, *p* = 0.03, Cohen’s *d* = -0.63).  
We can further ask whether it is the case whether our network entropy measure is approximating the entropy of hand-coded gestures. Figure 10 confirms that this is indeed the case, such that entropy increase based on human-coded information units is related to increase in entropy based on gesture network entropy.

Figure 11. Change in entropy in theme-level networks versus function-grouped networks  *Note Figure 11*. On the left panel, the average network entropy for the function-grouped gestures are plotted over the generations with red line showing the trend averaged over chain (other-colored lines). On the right panel this is shown for the gestures grouped by theme category. It can be seen that only the function-grouped gesture networks showed increased systematicity (lower entropy) over the generations.

We find that only functionally grouped tokens were minimizing entropy over the generations. Including generations for predicting function-level network entropy increased predictability as compared to a base model (random intercept chain, random slopes did not converge), chi-squared change (1) = 8.53, *p* = 0.00, model *R*-squared = 0.14, with generation relating to lower entropy *b* estimate = -0.0013, *t* ( 48.00 ) = -3.00, *p* = 0.00, Cohen’s *d* = -0.86).  
There was however no reliable decrease in entropy for the theme-level networks, chi-squared change (1) = 8.53, *p* = 0.00, model *R*-squared = 0.14.

## 3.2 Kinematic features

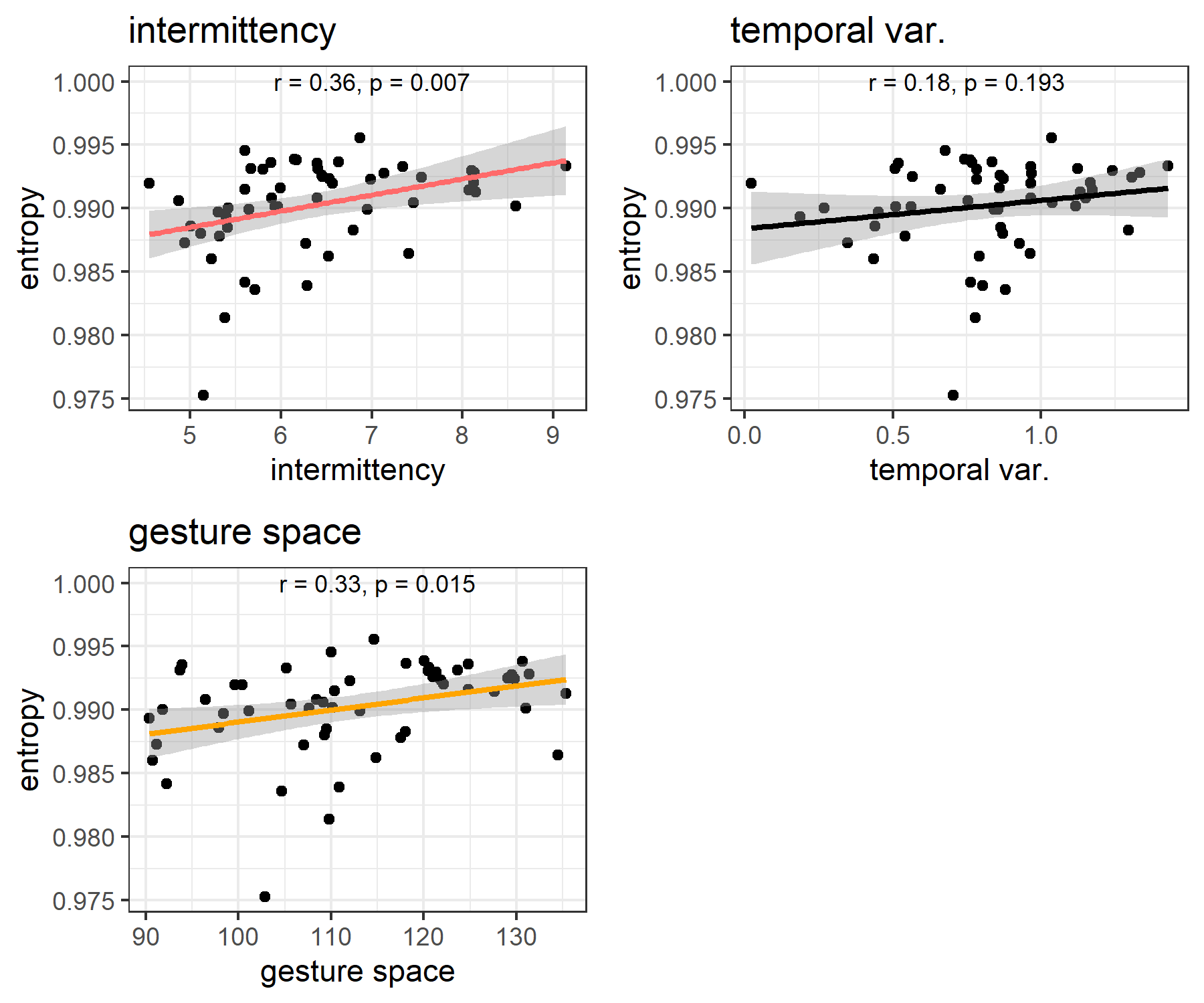
Next we performed mixed regression analysis for assessing potential kinematic changes as a function of generation, with random intercept for objects nested within chains (random slopes did not converge). See figure 12 for main results.

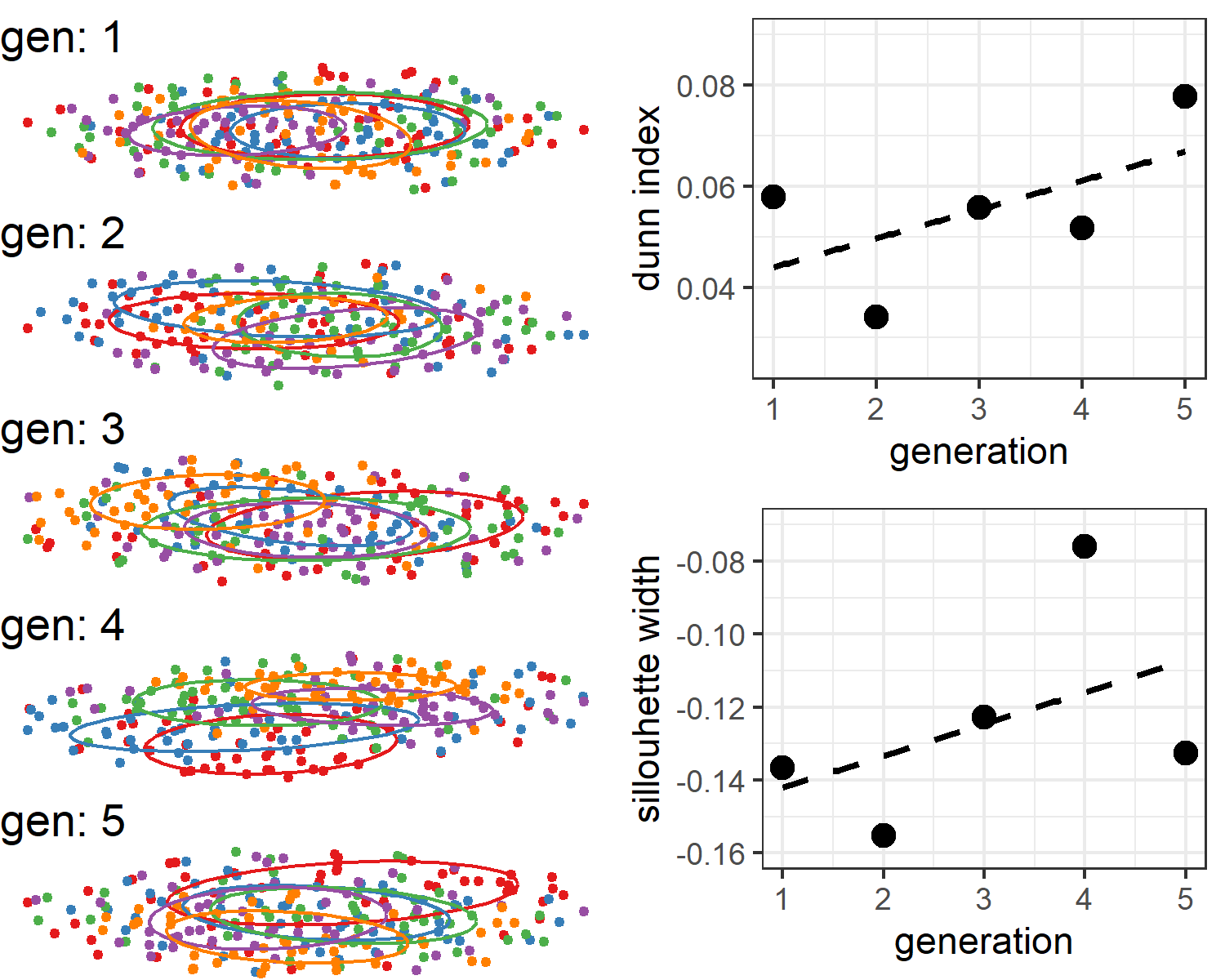
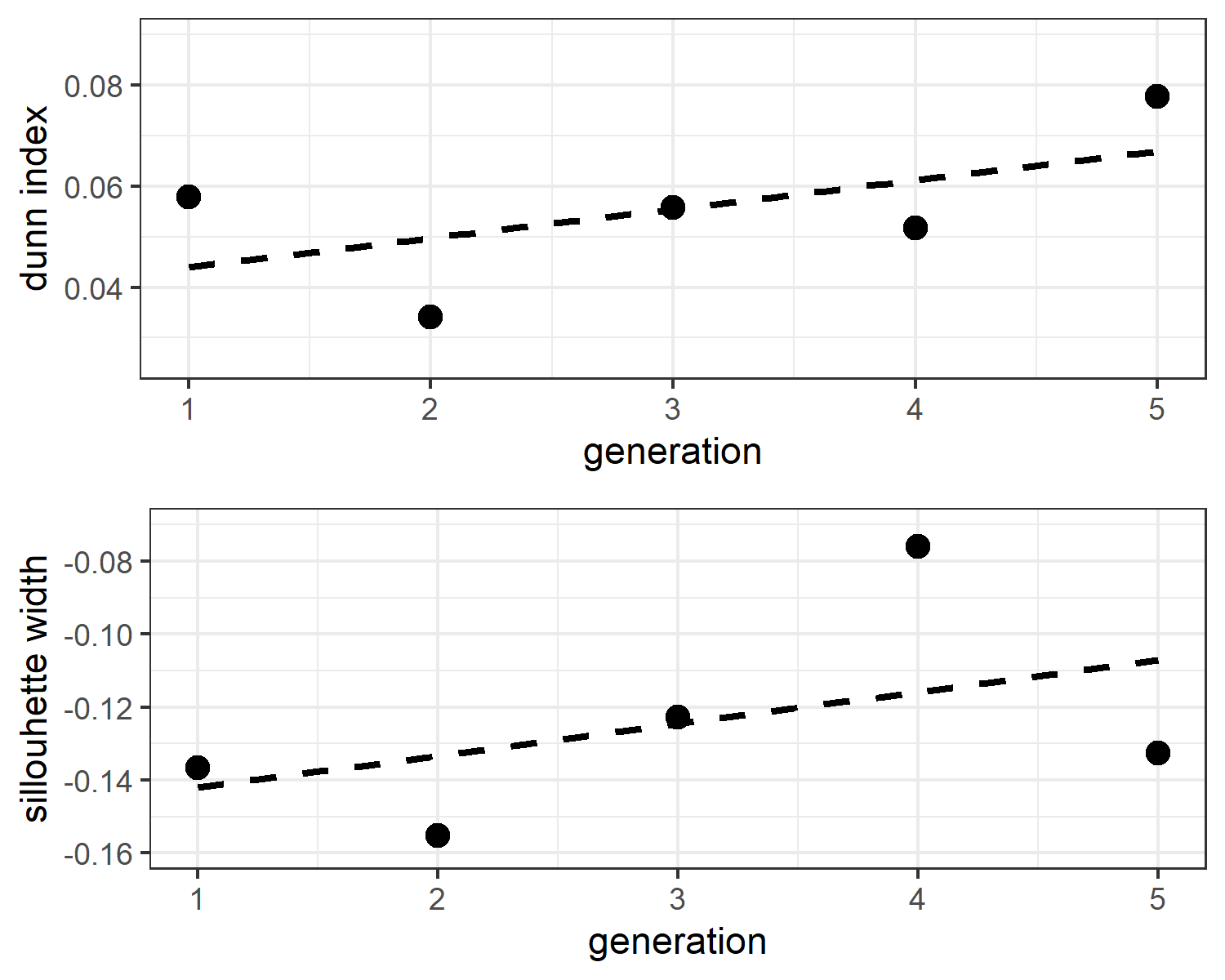
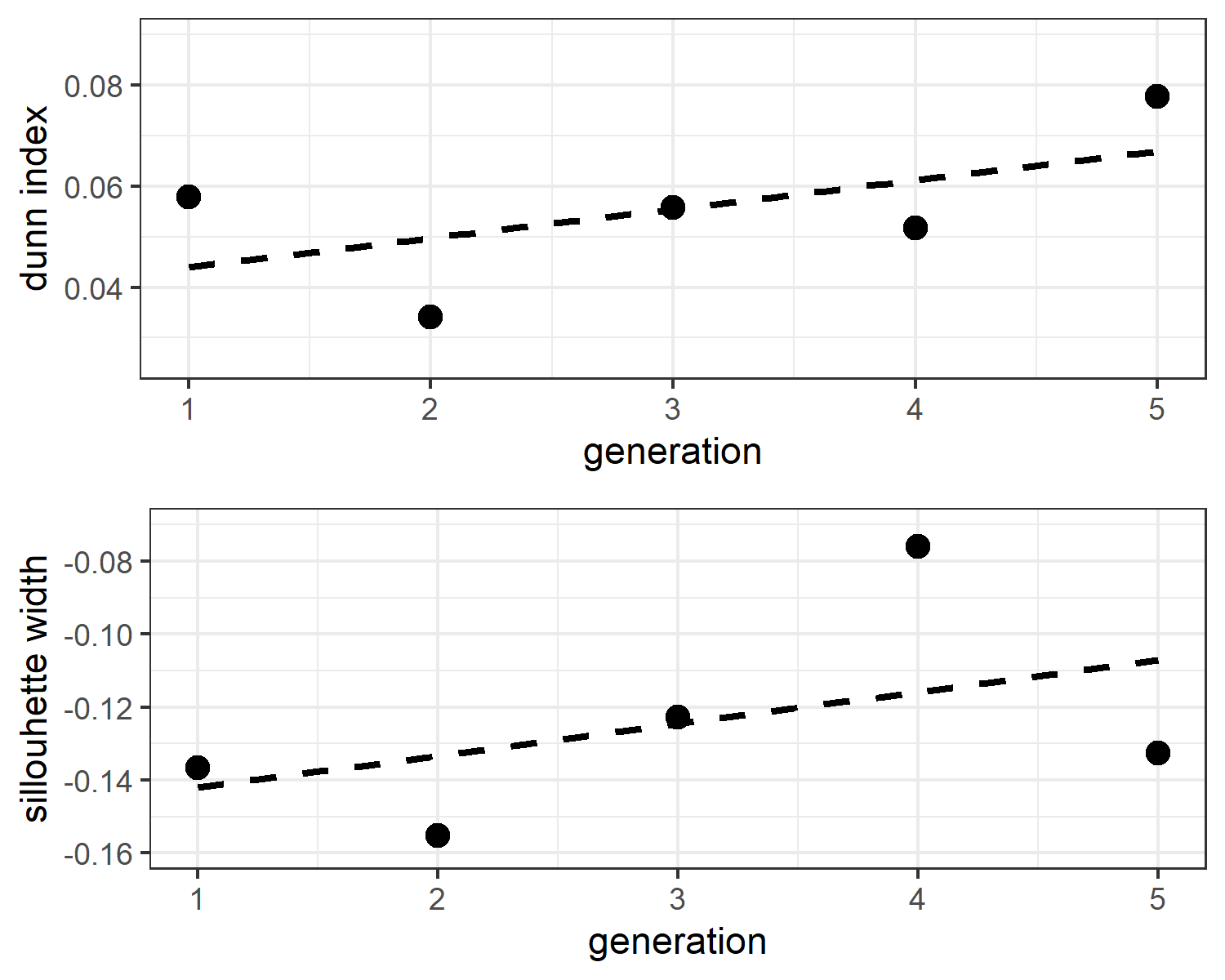
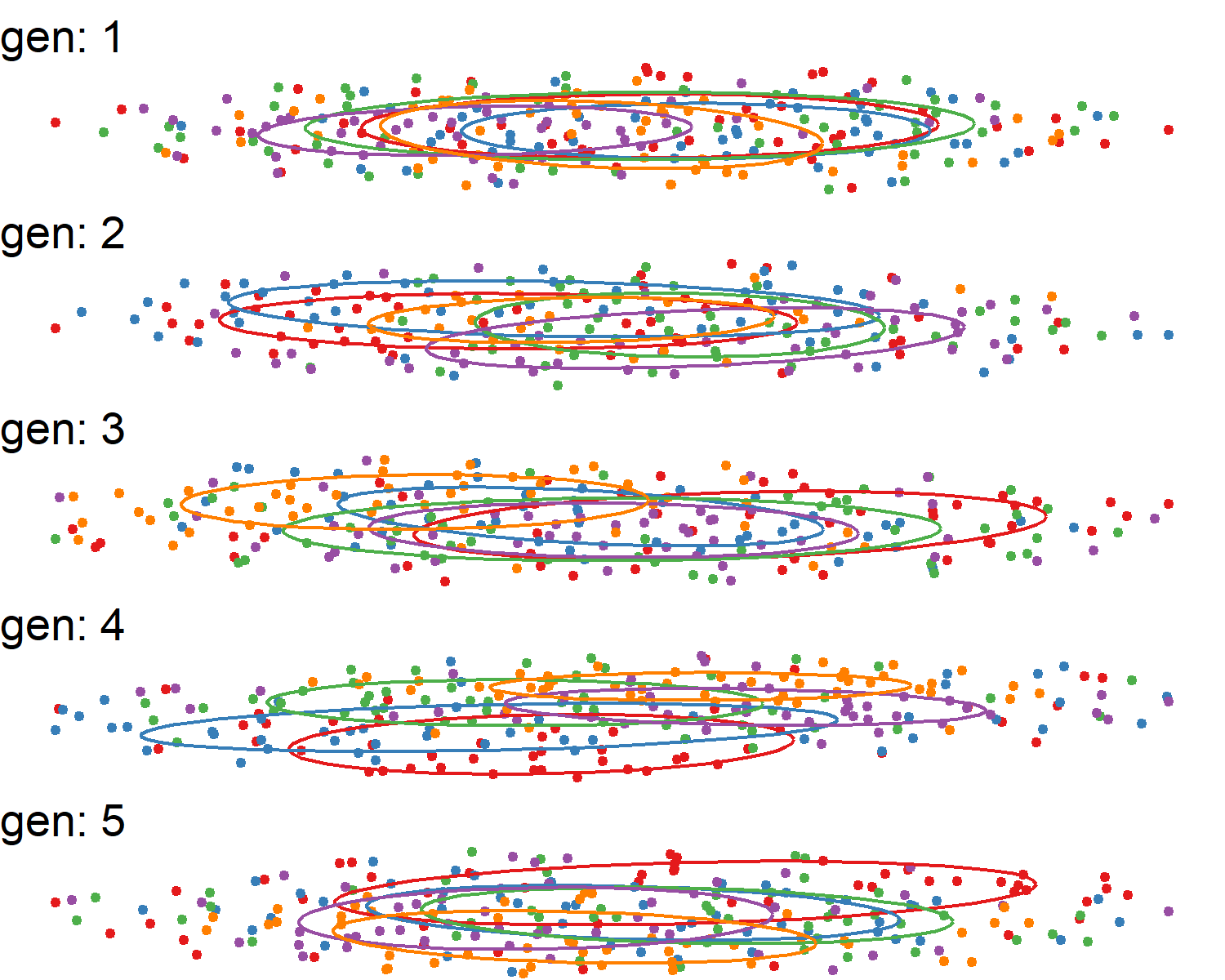
Figure 12. Change in kinematic properties over generations  
 *Note Figure 12*. Generation trends per chain are shown for intermittency, temporal variability and gesture space. Each observation indicates a communicative token, and these are spatially organized per their density distribution and colored by verb (green) or no verb (orange) (i.e., ‘action’ versus other function gestures). We can see that over the generations, movements become more smooth (lower intermittency score), with a more stable rhythm (lower temporal variability), and more minimized movements (smaller gesture space). Note, that temporal variability has lower data points as often the movement did not consist of more than 2 submovements. Thus, temporal variability indicates that *when there is a multi-segmented movement*, then such movements were more rhythmic.

Generations reliably predicted intermittency of the movements relative to a basemodel, chi-squared change (1) = 76.66, *p* <.001, model *R*-squared = 0.06. When adding verb as another predictor, this improved model fit for intermittency, chi-squared change (1) = 10.72, *p* = 0.00, model *R*-squared = 0.06. In this final model generation predicted lower intermittency score, *b* estimate = -0.2256, *t* ( 1,134.00 ) = -8.87, *p* < .001, Cohen’s *d* = -0.53). Silent gestures conveying verbs showed lower intermittency in general, *b* estimate = 0.3919, *t* ( 1,134.00 ) = 3.32, *p* < .001, Cohen’s *d* = 0.20). There were no interaction effects of generation and verb.  
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)). Adding verb or verb x generation to model temporal variability did not improve model fit. 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). Adding verb or verb x generation to model gesture space did not improve model fit. In conclusion, our kinematic results show all the hallmarks of increased communicative efficiency. Namely, gestures were on average smaller, less temporally variable, and less intermittent as the communicative system matured. Silent gestures that conveyed a verb were generally less intermittent, suggesting that they consist of smoother movement patterns.

## 3.3 Relations between Kinematic and network properties

Figure 11 contains the correlations of the relationships of kinematic properties (average per participant) and the network measures cluster persistence and entropy. Network entropy goes down as the average gesture space decreases, and the movement becomes less intermittent. This also comes at a trade-off, such that this simplification of kinematics also reduces differientiability of communicative tokens as shown by less stable clustering when gesture become smaller, less temporally variable, and less intermittent. Thus on the kinematic level there seems to be a general decrease of complexity which is further reflected on the level of the system as a whole as utterances become less *kinematically* differientable (less clustering) and more structured in their relations (lower entropy).

Figure 13. Relation between kinematic properties and network measures  *Note Figure 13*. Correlations are shown for each kinematic property averaged over all utterances and the concomittant network measure result. It can be seen that less intermittency, lower temporal variability, and smaller gesture spaces, relate to lower entropy and lower cluster persistence. This indicates that complexity in movement is cashed out in terms of systematicity and more homegeneous interrelationships (lower clustering) on the network level.



# 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, and then applied an analysis procedure to detect structural relations between gestural utterances (Pouw & Dixon, 2019). We found that communicative tokens showed higher systematicity at later generations, conceptually replicating results that were based on human coding of the gesture’s content (Motamedi et al., 2019). Indeed, gesture network entropy turned out to be a good approximation of entropy based on human coding of the gesture content. We further find that tokens were less stably differentiable on the form level as tokens have lower cluster persistence over the generations. Moreover, we found a decrease in entropy for the functional rather than the thematic dimension. While in the original study no increase in efficiency was found based on measuring gesture information units, we did detect increases of communicative efficiency for gesture kinematics. Over generations, gestures became less segmented (smoother), more rhythmic (if comprised of more than 3 submovements), and smaller. We also show that action gestures have a different kinematic quality as compared to non-action gestures, being more smooth (less intermittent) in their execution. Finally, we show that the decrease in kinematic complexity on the token level, predicts system-level changes of decreased entropy and decrease in clustering.  
That entropy decreased for gestures within the functional category at the level of kinematics, is consonant with the human coding findings of the original study and other related findings on sign languages showing regular employment of functional categories such as object- versus action distinctions (Padden et al., 2013). That gestures referring to actions are less intermittent as compared to non-action gestures in terms of their kinematics, conceptually replicates research based on human coding showing that action gestures often consist of a single segment (Ortega & Özyürek, 2020). In sum, our findings indicate that kinematics are revealing of the functional nature of gesture references, showing unique trajectories of change during iterated learning.  
A decrease in cluster persistence over generations here is likely to reflect the differentiability of communicative tokens, which as originally reported often showed iconic gestures at early stages in the iterations that were sometimes ambiguous in the theme category, and maximally differentiated from the other-themed gestures. For example, “arrest” and “police officer” could both contain a gesture that enacts the appliance of hand cuffs. Thus within themes there was clustering, but across themes there is differentiation. When gestures are disambiguated over the generations this will result in increased distances among the gestures within this category on the network level, i.e., leads to less clustering. While clusters became more unstable over the generations, the diversity of the interrelationships of the communicative tokens decreased (i.e., entropy decreased), and this is especially on the functional level. This suggests that there is a more consistent and thus homogeneous way in which the communicative system is organized, and the reorganization is caused by a reuse of gesture *across themes* and *within function*. That this increase in consistency is indeed a form of systematicity is fully supported by the detection of entropy decrease over the generations for communicative tokens grouped on the functional dimension (e.g., agent, action, location), but not the thematic dimension (e.g., justice, cooking).  
The kinematic findings suggest that the manual utterances simplify, in the way of reducing in size, 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, as making a minimal amount of smaller rhythmic movements reduces the degrees of freedom for articulation (Bernstein, 1967; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984). Moreover, this increased rhythmicity could also increase learnability and comprehensibility of the gesture, as we know from speech perception in noisy conditions that it is optimally perceived when speech is more rhythmic (Wang, Kong, Zhang, Wu, & Li, 2018).  
Interestingly, this reduction of degrees of freedom of the pronunciation, is precisely what one finds for novice learners of ASL. ASL learners have been found to spatially reduce their signs as they become more fluent (Lupton & Zelaznik, 1990; Wilbur, 1990). Moreover, a reduction in duration between the compounds of the signs have been observed during learning progression, where multicomponent component signs are increasingly performed as a single sign. In the present paradigm, there is a similar evolution of pronunciation, such that gestural multi-articulatory utterances acquire stable functional organizations across generations. Suboptimal organization of sub-movements will be filtered out as it were over the generations, and the temporally extended movement sequence becomes likely more coordinated whereby degrees of freedom are reduced by functioning as a single multi-aticulatory coordinative structure (Bernstein, 1967; Kelso, Tuller, & Harris, 1983), affecting for example gesture’s temporal variability and intermittency. That head movements improved differientation of real vs. falsely paired gestures in our analysis, further emphasizes that multiple articulators coordinated in the production of meaning in the current task. 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 which can attain unique linguistic functions which can then be combined in parallel into a single linguistically complex utterance.  
Note that our method allowed to account for the multiarticulatory nature of communication without formalized additional coding of the head movements, and we were able to quantify the unique communicative contributions of head and upper limb movements in the current paradigm. In this way, the current method is a bottom-up approach that will invite further investigation when needed. Our bottom-up approach further showed that gesture network entropy decreases alongside the entropy obtained from human coded content segmentation of the gestural utterance (Motamedi et al., 2019), suggesting that systematicity in form can be detected without the need for an apriori coding scheme. But our method as exposed here goes one step further. If gesture network analysis is complemented with kinematic feature analysis, it can be further assessed *what* is driving systematicity, providing insights on the evolution of the morphology of the silent gesture system. Coding schemes are notoriously difficult to formalize as any gesture researcher will confirm, and the current buttom-up method provides a formalized procedure for the detection of gesture evolution. As it is formalized and reproducible, the method is waiting to be applied to large datasets that are impractical to (completely) code by human annotators. An exciting avenue of further research is how different morphological evolutions can yield similar or different levels of systematicity at the gesture network level depending on different communicative constraints in vast populations. As such, we have shown that human-coded information units can be approximated from the kinematics (intermittency) and it can be assessed whether changes in such units are cashed out on the system level in the form of differientiability (clustering) and information compression (network entropy) - no human coding needed.  
There are two important caveats to the analyses presented here. First, in general, it is the case that kinematic analysis cannot say anything about the precise semiotic content that might evolve, and this is especially the case with incereasing ‘drifts towards the arbitrary’ (Tomasello, 2008). Although such drift might be detected via our network analysis, recognizing its possible semiotic content will always require extensive analysis (Sandler, 2018). However, since changes can be detected our analysis may invite human coding to be performed on a subset of the data which show promising changes over generations, inviting further necessary human interperation to understand what was driving such changes on the level of meaning.  
A second caveat is that there is a limitation to dynamic time warping. If we appreciate that compositionality increases as a communicative system matures, holistic gestures become segmented and the order of presentation of such segments might be varied. However, the dynamic time warping algorithm is sensitive to ordering and would judge two gestures containing identical segments in different orders as very different, while for a human coder the similarity between differently ordered segments might be transparent. Thus our 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 look for trajectory overlaps rather than ordering through time (Pouw & Dixon 2019), but such analysis goes beyond the current approach.  
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: It will reduce the amount of manual coding which is currently consuming many researchers’ time. It provides a much needed *multiscale* approach to how gestures evolve as communicative systems. The current method can further scale up the study of language evolution across modalities, as the kinematic analysis shown here functions much like an acoustic and articulatory analysis in speech. Beyond such promises, the current multiscale approach has shown that silent gesture kinematics evolve in structural ways during iterated learning.

# 5 References

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520. <https://doi.org/10.1016/j.jml.2007.02.004>

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

Bickerton, D. (2009). *Adam’s Tongue*. New York: Hill & Wang.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. In (pp. 7291–7299). Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Retrieved from <http://openaccess.thecvf.com/content_cvpr_2017/html/Cao_Realtime_Multi-Person_2D_CVPR_2017_paper.html>

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>

Deutscher, G. (2005). *The Unfolding of Language: An Evolutionary of Mankind’s Greatest Invention*. New York: Metropolitan Books.

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>

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., … Enfield, N. J. (2015). Universal Principles in the Repair of Communication Problems. *PLoS ONE*, *10*(9). <https://doi.org/10.1371/journal.pone.0136100>

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. (2016). *Natural causes of language: Frames, biases, and cultural transmission*. Berlin: Language Science Press. Retrieved from <https://langsci-press.org/catalog/book/48>

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3), 351–386. <https://doi.org/10.1111/j.1551-6709.2009.01090.x>

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press. Retrieved from <http://books.google.com?id=RProTk_Ag7gC>

Garrod, S., Fay, N., Lee, J., Oberlander, J., & Macleod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–987. <https://doi.org/10.1080/03640210701703659>

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., 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.

Kirby, S., & Christiansen, M. H. (2003). From Language Learning to Language Evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 272–294). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199244843.003.0015>

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. <https://doi.org/10.1073/pnas.0707835105>

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>

McNeilage, P. (2008). *The origin of speech*. New York: Oxford University Press.

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>

Ortega, G., & Özyürek, A. (2020). Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Language and Cognition*, *12*(1), 84–113. <https://doi.org/10.1017/langcog.2019.28>

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>

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

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>

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>

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1907), 20191262. <https://doi.org/10.1098/rspb.2019.1262>

Rączaszek-Leonardi, J., & Scott Kelso, 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>

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

Scott-phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*, 411–417. <https://doi.org/10.1016/j.tics.2010.06.006>

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>

Tomasello, M. (2008). *The origins of human communication*. Cambdride, MA: MIT press.

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. (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).

# 6 Supplemental

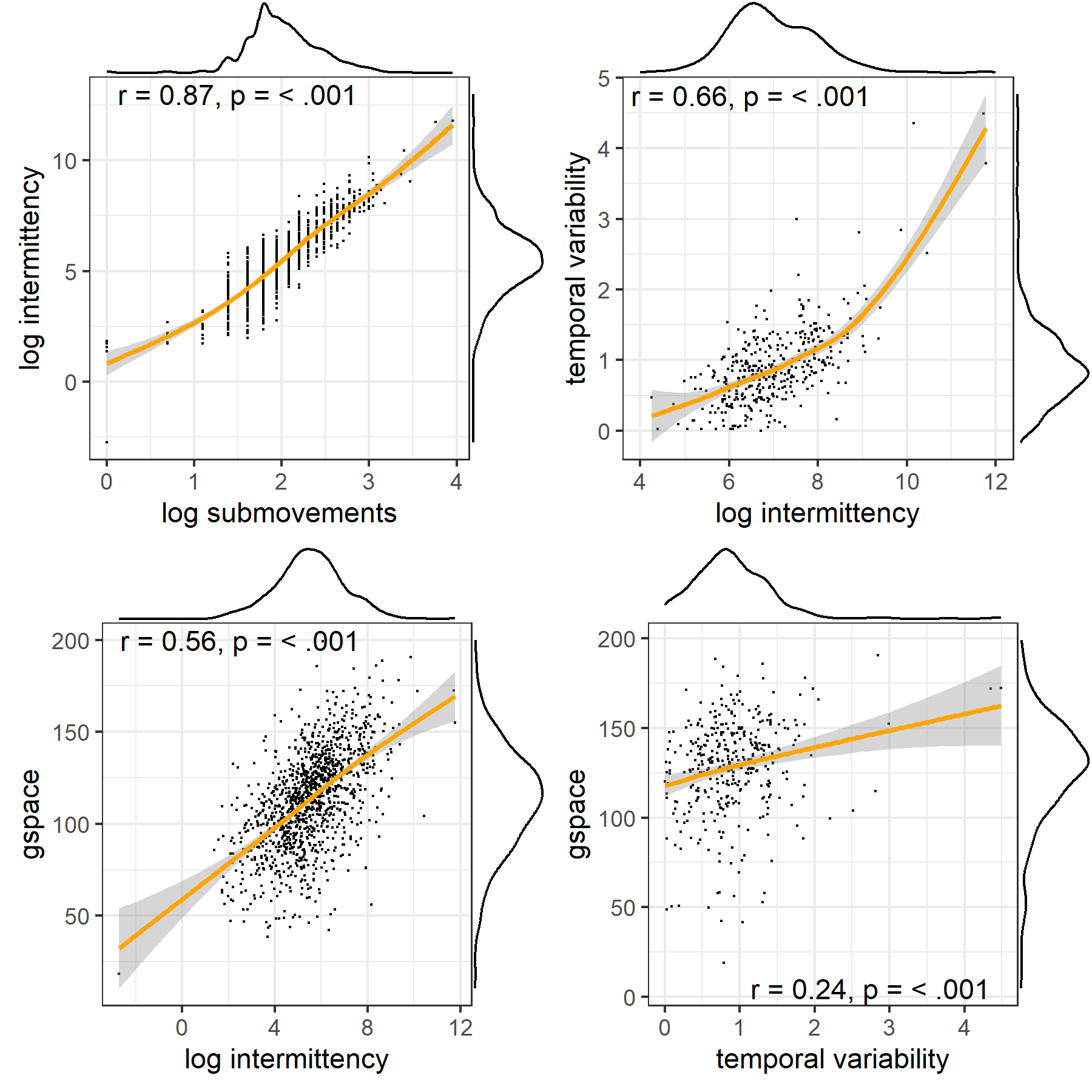
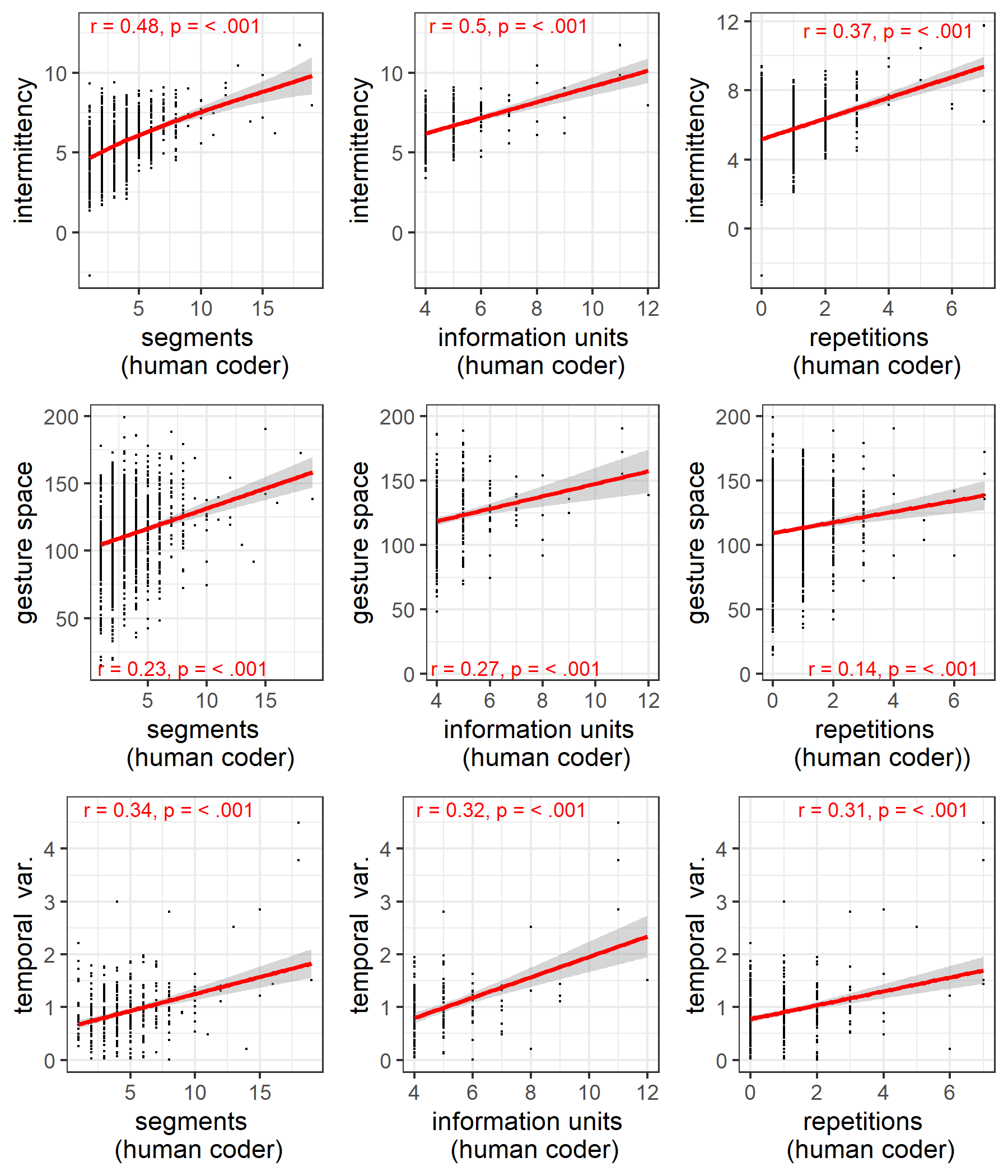
Figure S1. Correlations and distributions for kinematic measures per trial  *Note Figure S1*. Left upper panel, correlations and distributions are shown for intermittency and submovement. Given their high correlation we will use intermittency score for our final analysis. Other correlations are shown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