Conventionalisation and Discrimination as Competing Pressures on Continuous Speech-like Signals

Hannah Little[1,2] (hannah.little@mpi.nl), Kerem Eryılmaz[2] (kerem@ai.vub.ac.be)

& Bart de Boer[2] (bart@arti.vub.ac.be)

[1]Max Plank Institute for Psycholinguistics,

P.O. Box 310, 6500 AH Nijmegen, the Netherlands

[2]Artificial Intelligence Laboratory, Vrije Universiteit Brussel,

Pleinlaan 2, 1050 Brussel, Belgium

Abstract

Arbitrary communication systems can emerge from iconic beginnings through processes of conventionalisation via interaction. Here, we explore whether this process of conventionalisation occurs with continuous, auditory signals. We conducted an artificial signalling experiment. Participants either created signals for themselves, or for a partner in a communication game. We found no evidence that the speech-like signals in our experiment became less iconic or simpler through interaction. We hypothesise that the reason for our results is that when it is difficult to be iconic initially because of the constraints of the modality, then iconicity needs to emerge to enable grounding before conventionalisation can occur. Further, pressures for discrimination, caused by the expanding meaning space in our study, may cause more complexity to emerge, again as a result of the restrictive signalling modality. Our findings have possible implications for the processes of conventionalisation possible in signed and spoken languages, as the spoken modality is more restrictive than the manual modality.

Conventionalisation and Discrimination as Competing Pressures on Continuous Speech-like Signals

Introduction

Speech, on the whole, is arbitrary. That is, in modern language there is very little similarity between spoken words and the meanings they refer to. However, having signals which are similar to their referents in some way (iconicity) is one way in which language could have initially bootstrapped itself as a communication system (Imai & Kita, 2014). If a signal is similar to its referent in some way, it will be easier for language users to establish a signal-meaning mapping. However, there is very little direct evidence available from real world languages about how language initially bootstrapped itself, especially spoken languages. As a result, experimental studies have been used by researchers in the field of language evolution to investigate the effects of interaction and transmission on levels of iconicity and symbolism in signals. Specifically, studies have concentrated on how we could have got from iconic beginnings to an arbitrary system via processes of conventionalisation.

One of the main methods for investigating the process of conventionalisation has been the field of experimental semiotics (see Galantucci & Garrod, 2011, for a review). This started as far back as Brennan and Clark (1996), where participants communicated different concepts using tangrams. Tangrams are arrangements made up from 7 flat shapes. They found that after repeated interactions, the tangram arrangements became more simplified as participants started to use elements of the original tangram arrangements as “short-hand". This simplification of originally iconic forms, leading to a loss in iconicity, is the hallmark of conventionalisation as we use it throughout the rest of this paper.

Since Brennan and Clark (1996), Garrod, Fay, Lee, Oberlander, and MacLeod (2007) have explored how iconic signals evolve into symbolic representations using a pictionary-style task in different conditions. Garrod et al. (2007) had 3 conditions. In one condition, one participant repeatedly drew items for an imaginary audience (no feedback). In another, one participant drew items but were given feedback from a partner. In the final condition, two participants took it in turns to draw items for each other with ongoing feedback. The study measured complexity in the images throughout the task, as well as the levels of iconicity in the drawings. They measured iconicity with the rate at which naïve participants could match the drawings with their intended referents after the experiment. Garrod et al. (2007) showed that knowledge of early interactions in the communication condition of the experiment improved naïve participant's ability to match drawings with their referents, indicating that the images were becoming less iconic. Getting naïve participants to match signs with referents is now a common method used in experimental semiotics to measure iconicity. If naïve participants can pair signals with their intended meanings, then those signals can be said to be iconic. Garrod et al. (2007) also found that complexity in the images dropped throughout the communication condition, as it did in Brennan and Clark (1996). However, in the individual condition, with no communication partner, the drawings increased in complexity.

Other studies which used graphical signs to investigate conventionalisation include Theisen, Oberlander, and Kirby (2010), which also used a pictionary style paradigm in a communication task. They showed that over the course of the communication game, drawings became less iconic. One of the contributing factors to minimise production effort in this experiment was an incentive for participants to have as many successful communicative interactions as possible within a constrained time period. A slightly different approach was demonstrated by Caldwell and Smith (2012), which had “replacement microsocieties", where they had a constant turnover of naïve participants who contributed to signs becoming simpler and more abstract. One of the driving forces for signs becoming simpler in this experiment was that participants could interrupt the production of a signal once they were sure what it was, meaning signals never had to be more complex than they needed to be. Concurrent feedback, such as interruption, was also found to drive conventionalisation in conditions in Healey, Swoboda, Umata, and King (2007) and Garrod et al. (2007).

There have also been several studies which have used gestural experiments to investigate whether conventionalisation happens through interaction to get from iconic pantomime-like gestures to more arbitrary language-like symbolic gestures. Namboodiripad, Lenzen, Lepic, and Verhoef (2016) used a communication game in the lab to get participants to repeatedly communicate scenes to one another and were able to measure hallmarks of conventionalisation over the course of the experiment. Duration of gestures and the size of the space used for the gestures was reduced, as was the amount of complexity within a gesture. Motamedi, Schouwstra, Smith, and Kirby (2016) also investigated conventionalisation in silent gesture, but focused on the effect of transmission rather than interaction, looking at how signs changed in an iterated transmission chain, where participants’ signs were learnt from those output of a previous participant pair. This study found that gestures developed from pantomimes to less complex, more arbitrary signs.

Real world data can also contribute to our knowledge of conventionalisation processes. There is diachronic evidence of some signs in American Sign Language (ASL) losing complexity and iconicity (Schlehofer, 2016). Evidence from younger, emerging sign languages, such as Al-Sayyid Bedouin Sign Language (ABSL) indicates that the emergence of the first combinatorial phonology-like elements in the language may be the result of a loss of iconicity in some signals as a result of conventionalisation (Sandler, Aronoff, Meir, & Padden, 2011).

Combinatorial structure (structure where meaningless building blocks combine to make meaningful units), has been hypothesised to have emerged in sign languages as an alternative strategy to iconicity (Goldin-Meadow & McNeill, 1999). Spoken language has high levels of combinatorial structure because the spoken modality is less able to iconically represent meanings than the sign modality. In emerging sign languages, Goldin-Meadow and McNeill (1999) propose that once an element of a signal ceases to be interpreted as iconic, as a result of conventionalisation, then it opens itself up to be reanalysed as a meaningless building block which can then be reused as combinatorial units. Several studies since have used continuous signal-space paradigms, such as that used in Roberts and Galantucci (2012), to look at whether iconicity hinders the emergence of combinatorial structure in signals. Roberts, Lewandowski, and Galantucci (2015) used a communication task where it was either easy or difficult to create iconic signals for meanings. They found that when it was more difficult to be iconic, then combinatorial structure was more likely to appear. Verhoef, Kirby, and Boer (2015) carried out an experiment which investigated signals produced using slide whistles. They used an iterated learning paradigm where signals from participants were fed to other participants in a transmission chain. Signals and meanings were either kept matched to one another in one condition (facilitating iconic mappings), or in the other condition, meanings and signals were paired randomly between each generation in the experiment. This study found that the emergence of structure was slowed down when the signal-meaning pairs were kept stable, indicating that the iconicity in this condition was inhibiting the emergence of combinatorial structure. However, neither of these experiments looked explicitly at the process of conventionalisation across time, instead opting to have different conditions which either facilitated or inhibited the use of more iconic signals.

Another hypothesis for the emergence of combinatorial structure is that of Hockett (1960). He was the first to hypothesise that combinatorial structure emerged as a way to deal with pressures for discrimination caused by larger meaning spaces. This hypothesis was also tested by Roberts and Galantucci (2012) who investigated whether signal repertoires for bigger meaning spaces had more combinatorial structure, though their results were inconclusive.

Our Study

In the current study, we compare signals produced in an individual condition (where individuals both produce and recognise their own signals), with a communication condition between two individuals. We then have naïve listeners match signals from both conditions with referents from both the beginning and end of the experiment in order to see how signals changed over the course of the experiment.

The signals used in our experiment are more analogous to speech signals than the drawings in Garrod et al. (2007), or signals made from pre-discretised units, as in studies such as Kirby, Cornish, and Smith (2008). Our signals are continuous, auditory and make iconicity more difficult than it is with graphical representations. However, the signals still remain non-linguistic enough to inhibit interference from pre-existing linguistic knowledge.

In both conditions in our experiment, we have a growing meaning space, allowing for investigation into the effect of discrimination pressures as the meaning space expands, both on iconicity and structure in signals.

Hypotheses

One of the main hypotheses investigated in Garrod et al. (2007), was whether complex iconic representations become more abstract symbolic representations through a process of repetition, or whether interaction was also a necessary driving force in this process of conventionalisation. We are interested in whether this process of conventionalisation also happens in more speech-like signals.

If processes of conventionalisation happen in our continuous auditory signals in the same way as they do with pictorial representations, following from Garrod et al. (2007), we expect to see two things:

1. In the communication condition, signals will lose complexity throughout the experiment. In contrast, in the individual condition, signals will gain complexity.

2. In the communication condition, signals will lose iconicity and in the individual condition iconicity will be retained.

We are also interested in how knowledge of another person in the experiment will influence what the signals look like. We hypothesise that the knowledge that a signal is meant for someone else may drive signals to be more iconic, in order to aid bootstrapping, before conventionalisation can happen. This difference (if present) will be evident in the difference in iconicity between conditions at the beginning of the experiment.

We are also interested in whether combinatorial-like structure emerges and whether this will correlate with a loss in iconicity due to conventionalisation, as hypothesised by Goldin-Meadow and McNeill (1999).

Further, we are interested in the hypothesis of Hockett (1960), that signals may adopt combinatorial-like structure to deal with pressures for discrimination caused by our growing meaning space. We might expect the signals to grow in complexity to assist with the task of discrimination. This effect could be negated by the process of conventionalisation, or occur in spite of it, or possibly in tandem with it.

Experiment

Signals

Participants created signals using a “Leap Motion" hand-tracking device: an infrared sensor designed to detect hand position and motion (Eryılmaz & Little, 2016). Participant’s hand position was translated to the pitch of audio signals. Moving their hand to the left would make the signal lower, moving their hand right would make the signal higher. It was not possible to make pauses in a signal. Signals had no time limit. All participants were given a demonstration of how the sensor worked before they started the experiment, as well as time to use it themselves, to get used to the mapping between hand-position and auditory feedback. The mapping between hand position and auditory feedback was not linear. The auditory tone generated was an exponential function of the x-coordinate of the hand position in the space above the sensor. The function was exponential because of the non-linear way humans perceive pitch. The x-coordinates ranged from -250-250 (with 0 being the centre point of the signal space). These coordinates were transformed to pitch using the following formula:

$$frequency = 110 \times 3^{\frac{(|x+200|)}{200}}$$

Meanings

The meaning space was constructed to have no internal structure. The meanings all had three features: shape, colour and texture. No two meanings had any features in common. For example, in figure 1, only the left shape had the features blue, circle and stripey, and only the cross had the features grey, cross and wavy lines. There were 15 meanings in the experiment (see figure 2).

Conditions

Participants were assigned to one of two conditions; an individual condition or a communication condition. In the individual condition, participants both produced and recognised their own signals. In the communication condition, participants took it in turn to produce and recognise the signals of a partner. It is important that we had signal creation and recognition within both conditions, allowing for; i) comparable measures of recognition accuracy from within the experiment, and ii) a pressure for expressivity in both conditions allowing for isolation of effects caused by the communication of two people, rather than the process of communication itself (as individuals are effectively communicating with themselves in the individual condition).

Individual Condition

Participants. 24 participants (17 female, 7 male, average age 21 ± 1.3) took part in the individual condition and were paid €5 for the 30 minutes it took to complete the experiment. Participants were recruited at the Vrije Universiteit Brussel.

Procedure. Participants were given clear instructions about the structure of the experiment. They were explicitly told how many phases there were, and how the phases were structured. They were also told how to use the leap motion by simply moving their hand either left or right to manipulate the pitch of the signal. They got to try this out before the experiment began. Participants knew from the beginning that they would have to recognise their own signals in each round.

Phases. Participants created signals in three phases (see figure 2). In the first phase, they created signals for 5 meanings, chosen at random from the pool of 15. In phase 2, they created signals for all of the meanings they had already seen, plus 5 more, making 10 in total. In phase 3 they created signals for all 15 meanings.

Signal Creation Task. Before the signal creation task, participants were told that they would see images which they need to create signals for. They were explicitly told they should make sure they remember the signals as they would be asked to recognise the signals during the experiment. This introduction screen also displayed the whole meaning space for that phase, so participants knew which meanings were in a phase before they began creating signals.

Meanings were presented one after another in a random order and participants created a signal for each one by pressing a “record" button to start, and a “stop" button to finish. Participants could play signals back and rerecord them if they were not happy.

Signal Recognition Task. Once participants had created signals for all meanings within a phase, they were given a signal recognition task. They heard each of their signals in a random order, one after the other, and were asked to identify the meaning it referred to from an array of 4 choices. The array included the correct meaning, and 3 other meanings taken randomly from the subset of meanings used within the current phase. Participants were given feedback on the correct answer immediately after each response.

Practice and Experimental Round. Participants completed the signal creation task and the signal recognition task twice for each phase. The first time was framed as a “practice round", and existed so that the participant could get used to the structure of the experiment and how to use the apparatus. Only data from the experimental round was used in the analysis of the experiment.

Post-experimental questionnaire. After the experiment, participants completed a post-experimental questionnaire. It asked about the specific strategies participants used to generate signals, and whether they felt their strategies changed at all throughout the course of the experiment.

Communication Condition

Participants. 32 participants (27 female, 5 male, average age 20.9 ± 2.8) took part in 16 pairs in the communication condition. In this condition participants were paid €10 for 1 hour. Participants were recruited at the Vrije Universiteit Brussel.

Procedure

Again, participants were given clear instructions about the structure of the experiment, but were not given a detailed explanation about the mathematics of when and how the meaning space expanded (see below). A detailed explanation would not have served their success in the experiment, but may have confused or distracted them from the simple goal of communication. They were told they would be playing a communication game. Again, they were told how to use the sensor and given an opportunity to practice making signals before the experiment began. They were told how a turn worked. They were told that they would be given feedback about their success after each turn, and that they would take it in turns to produce signals. They were told that if they had not finished the experiment after 50 minutes, then the experiment would automatically end.

Participants also knew that the experiment would progress more quickly the more successful they were (the specifics of this mechanism are explained below). Participants were also given an incentive to try to finish the experiment quickly. They were told that the pair of participants who do the experiment the fastest would win a €20 voucher.

2 participants took it in turns to produce and receive signals with the producer creating a signal for a meaning and the receiver choosing from an array of up to four meanings, as in the individual condition. Both participants were given feedback after every interaction about whether their communication was successful, as well as feedback about both the meaning the producer was communicating, and the meaning the receiver chose.

As in the individual condition, the communication condition also had an expanding meaning space by phase. However, the meaning space only expanded by 2 meanings at a time (rather than 5 in the individual condition) and the experiment only continued to the next phase once the participants had agreed on signals for existing meanings. Ideally, the meaning space should have expanded at the same rate as the individual condition. However, participants found the communication game much more difficult than we had anticipated when giving the participants the meanings in batches of 5. As a result, we designed a system where the meaning space expanded in line with their success in the experiment, in order to not overwhelm the participants with too many meanings at once. This setup ensured that meanings were seen potentially more (or less) than in the individual condition depending on how many times a meaning got randomly chosen. However, overall frequencies were comparable and meanings introduced earlier were seen more times than later ones, as in the individual condition. Bigger differences occurred if participants were particularly bad at the communication task, then they were given the same meanings many more times than in the individual condition.

Participants started with 2 meanings, chosen at random. The array in the recognition task was constrained to these 2 possible meanings at the beginning of the experiment. Once meanings had been communicated correctly twice in a row, they were considered “established" meanings. If an established meaning was communicated incorrectly, it would lose its established status. Once all meanings in a phase were established, then the meaning space expanded by 2 more meanings, starting a new phase. Since there were 15 meanings, the meaning space expanded only 7 times (the last time by only one meaning), making 8 phases in total.

At first, which meaning the pair were to communicate in each interaction was presented at random. However, once the meaning space expanded once, meanings were chosen for interactions with a probability determined by whether it was an established meaning or not. Meanings were chosen with a 45% probability if they were established, and the remaining 55% of the time the meanings were either newly introduced meanings, or meanings which had recently been communicated unsuccessfully. This mechanism was in place because if all meanings had the same probability of appearing throughout the experiment, the experiment would take far too long. Unestablished meanings needed to have a reasonable frequency in order to become established so that the experiment could progress.

Once all meanings were established, the experiment finished automatically. If participants did not achieve established signals for all 15 meanings before 50 minutes, they were stopped and their interactions and signals were recorded up until that point.

The signal data used in the analysis of this experiment was taken from signals once they had become established, in order to make them more comparable with the signals created in the experimental rounds in the individual condition.

Participants completed a post-experimental questionnaire, as in the individual condition.

Analysis of Signals

Signals for analysis from the individual experiment were either taken from the first phase (the first 5 signals produced) or the third phase (the last instance of all 15 signals), so we could measure how iconicity was affected by repetition of signals throughout the experiment.

Signals from the communication experiment were either from the first phase (for the first 2 meanings after they had been communicated correctly twice in a row) or they were the last successful instance of signals produced in the last phase of the experiment that a pair saw, which was dependent on how well they did in the experiment. Some pairs got to later phases than others as they were more successful at producing established meanings. This data is presented below.

Signal Measures. Garrod et al. (2007) measured **complexity** by calculating the amount of ink which was used in an image. In our study, we've made some effort to create a comparable complexity measure: the amount of "auditory ink" used in signals. This has been calculated by the duration of signals and the amount of movement in signals. This does mean that signals using similar movements but using more or the meaning space will be judged as more "complex". However, the amount of the signal space used has also been used to measure signs of conventionalisation in silent gesture studies such as Namboodiripad et al. (2016). The amount of movement in a signal was calculated by measuring how much of the signal space had been used in the signals of one participant. We measured this using the standard deviation of the trajectory of x-axis coordinates in each signal. The articulation space which could be utilised was 500 coordinates across. Each signal's data was made up of a list of coordinates which could be used to regenerate that signal. Using this information we could calculate the mean coordinate of a signal (mapping on to a signal's mean pitch) and also the amount the signal deviated from the mean. If a participant uses more of the signal space, their signals' coordinate standard deviations will be bigger. The duration of signals was simply measured using the number of data frames in a signal, which we converted to seconds for the purposes of presenting the results.

Further to the above, we also measured the predictability (or entropy) of signals based on the rest of a signal repertoire. This measure is similar to compressibility measures (e.g. Ehret & Szmrecsanyi, 2011) in that it is affected by repeated patterns in signal repertoires or static states, as these will make signals more predictable.

We measured the **predictability** within signals using the conditional probabilities of points within the signal trajectories. The points of a trajectory were quantised signal coordinates derived using a k-means algorithm ($k = 150$). Such a high value for k ensured that we represented our very fine-grained data effectively. The k-means algorithm clustered points in the trajectories into a time series of integer values representing a participant's entire repertoire of signals. With this, we estimated the marginal probability distribution of the points on each quantised trajectory and used these to calculate the conditional probabilities of individual points, and finally, the joint probability of whole signal trajectories by taking the negative logarithm of the product of first order conditional probabilities of the points on the trajectory.

This predictability measure allows us to measure structure at the level of a repertoire. In real language, combinatorial structure is not measurable at the level of one word. For example, if you only have the word "cat" you cannot know if any of its units exist in any other context, so you do not know if they are combinatorial. Measures of entropy or compression which measure each signal individually cannot tell us anything about the combinatorial structure of a signal repertoire, though can be informative about general complexity.

In order to measure **iconicity**, we did an online playback experiment with the signal data produced in both conditions. We asked naïve listeners to match the signals with the meaning they felt the signal most represented. 391 naïve listeners were recruited on social media. Each participant was asked to listen to 15 mp3 signals each and asked to choose from an array of 4 possible meanings for each signal. Some participants matched fewer than 15 because of experimenter error. Their data was still used in the analysis.

Results

The results will be presented in two parts: those results pertaining to the signals, followed by the results pertaining to the signal recognition tasks, both within the experiment, and after the experiment by naïve listeners.

Signals

Movement in signals. To investigate what affected the amount of movement in signals, we conducted a linear mixed effects analysis, with standard deviation as the dependent variable and how early in the experiment a signal was produced (phase number) and condition as fixed effects. We had participant number and meaning as random effects, as well as by-participant and by-item random slopes for the effect of both time produced and condition which were correlated with the intercepts. We then conducted likelihood ratio tests of our model against a null model without the effect in question (but with the same random slopes) in order to obtain p-values. We found that condition affected the amount of movement in a signal ($\chi^2(1) = 6.9$, $p = 0.009$), with signals from the individual condition having standard deviations which were lower (by on average 21.7mm), indicating less movement in the signals (see figure 3). However, how early in the experiment participants produced the signals did not significantly affect movement in the signals ($\chi^2(1) = 0.13$, $p = 0.25$). We also tested to see if there was an interaction between condition and time produced by comparing models with and without the interaction ($\chi^2(1) = 1.769$, $p = 0.18$).

Length of signals. We conducted a similar linear mixed effects analysis as with the standard deviation values above, to investigate the length of signals with the same random and fixed effects and random slopes. Signals produced in the communication condition were longer than in the individual condition, though this effect was not significant ($\chi^2(1) = 0.4$, $p = 0.52$). However, the time produced (phase number) did have a significant effect on the duration of signals ($\chi^2(1) = 4.4$, $p = 0.03$). As can be seen in figure 4, the duration went up throughout the experiment in both conditions, though this was more marked in the individual condition.

Predictability of signals. We conducted a similar linear mixed effects analysis as above looking at predictability with the same random and fixed effects and random slopes. Condition did not have an effect on the amount of predictability within signals ($\chi^2(1) = 0.02$, $p = 0.88$). There was an overall significant trend of production time ($\chi^2(1) = 5.53$, $p < 0.02$) though figure 5 indicates this may be primarily driven by the individual condition. However, there was no interaction between condition and production time ($\chi^2(1) = 1.44$, $p = 0.23$).

Signal Recognition

Recognition of signals within the experiment

We conducted a linear mixed effects analysis to look at participant success throughout the experiments, with time produced and which experiment signals were produced in as fixed effects. We had meaning and participant (or pair) number as a random effect, as well as by-meaning random slopes for the effect of time produced. As above, we then conducted likelihood ratio tests of our model against a null models to obtain p-values. Which experiment signals were created in had a significant effect on participant success within the experiment ($\chi^2(1) = 7.8$, $p = 0.005$), with participants being better in the individual experiment (85.6% correct) than in the communication experiment (74.4% correct). There was no significant effect of time produced on success during the experiment ($\chi^2(1) = 0.35$, $p = 0.55$). However, there was a significant interaction between experiment and time produced ($\chi^2(1) = 5$, $p = 0.02$). As can be seen in figure 6, in the individual experiment, participants got slightly better throughout the experiment. In the communication experiment, participants got worse.

Another measure of success within the communication condition was how far participants got before their time ran out. As explained in the methods, whether participants got to the next phase was dictated by whether they had managed to establish signals for all of the meanings which were currently in the meaning space. As one would expect, some pairs were much better at the task than others, with some pairs only reaching the second phase of the experiment (4 meanings), and others doing much better (success of all pairs can be seen in figure 7). No pair managed to establish signals for all 15 meanings, thus, nobody finished the experiment. As a result of this, the signals used in the playback experiment were taken from signals at the end of the experiment no matter where they got to in the communication condition, rather than using signals from specific phases.

Recognition of signals by naïve listeners

We conducted a linear mixed effects analysis, with time produced (early or late) and condition as fixed effects. We had meaning as a random effect, as well as by-meaning random slopes for the effect of time of production and condition. Again, we conducted likelihood ratio tests of our model against a null model. Condition did not affect the amount of iconicity in the signals ($\chi^2(1) = 0.1$, $p = 0.74$), with overall levels of matching nearly exactly the same (around 35% in both conditions). How early in the experiment participants produced the signals also did not significantly affect iconicity ($\chi^2(1) = 2.3$, $p = 0.13$). However, there was a significant interaction between condition and time produced ($\chi^2(1) = 5.9$, $p = 0.015$). As can be seen in figure 8, naïve listeners were much better at matching signals with their intended referents which were produced later in the experiment in the communication condition. However, in the individual condition, the signals went down in their iconicity, though this difference was much less marked than in the communication condition.

We were also able to measure the iconicity of signals for specific meanings. Figure 9 shows the iconicity of each signal as measured using naïve listeners. Some meanings lend themselves to iconicity better than others. The upwards pointing arrow is particular strong in its iconicity, almost certainly because having a signal with rising pitch is an easy way to represent this in the paradigm. Signals for pointy images were also easy to recognise, though some participants in the communication condition did report having trouble differentiating the signals of their partners' for these meanings.

Post-experimental questionnaire. The questionnaire revealed that nearly all participants attempted to use iconic strategies throughout the experiment in both conditions. They were more likely to try and use shape than any other feature to identify signals.

Discussion and Further Work

In our experiment, we measured complexity in signals to give us some sense of whether signals were becoming more simplified throughout the experiment, from more complex iconic representations to a more abstract symbolic representation as we hypothesised according to the results of Garrod et al. (2007).

We have used several measures to quantify complexity in our signals (movement, duration and predictability). Using these measures, we found that signals were less complex in the individual condition than in the communication condition. This is in contrast to findings from Garrod et al. (2007), who found that pictures produced in their individual condition stayed complex throughout the experiment, and pictures produced in the communication condition reduced in complexity throughout, resulting in the images in the communication condition overall to be much less complex, the opposite of our finding. In our experiment, we found no effect of signals becoming less complex over time in the communication condition, an effect that is likely to be due to the differences between our signalling paradigm and that of Garrod et al. (2007).

Signals in our paradigm are much more constrained in the forms they can take, which may mean they need to grow in complexity simply in order to differentiate between different meanings in the experiment as the meaning space expands. Under the hypothesis of Hockett (1960), that a growing meaning space will elicit combinatorial structure because of crowding in the signal space, we might not expect the signals to become simpler in either the individual or communication conditions as further complexity is beneficial for the task of discrimination. The reason the drawings in the communication condition in Garrod et al. (2007) dropped in their complexity was possibly because their communication modality (drawing) was so much more flexible than our paradigm, allowing for more complexity as a starting point. With the signal-space being so much more restricted with the Leap Motion signals, participants started simple, and ran out of ways to generate distinctions between signals quite quickly. This may have implications for the processes of conventionalisation (or the emergence of combinatorial structure) between languages in the real world, as the signed modality is arguably much more flexible than the spoken modality. Indeed, in emerging sign languages, such as ABSL, we can observe a delay in the emergence of combinatorial structure (Sandler et al., 2011), which is possibly because the flexibility of the modality does not immediately produce the pressure described by Hockett (1960). This pressure for discrimination (or expressivity) is also often cited as important factor in the emergence of structure in artificial language experiments which use pre-discretised building blocks to form signals (e.g. in Carr, Smith, Cornish, & Kirby, 2016; Kirby et al., 2008; Kirby, Tamariz, Cornish, & Smith, 2015)).

Further, in our experiment we found that signals became more complex later in the experiment in the individual condition, which is in line with the findings from Garrod et al. (2007). However, Garrod et al. (2007) hypothesise that their result is because, in the absence of feedback, participants encode more features in their signals later in the experiment as they think of more things they can include about the meanings they are communicating. The opportunity for this to happen in the current experiment was relatively limited, as the meaning space was not so complex. Further, in post-experimental questionnaires participants usually only describe trying to encode one feature of the meanings (mostly shape). As a result, the pressure for discrimination from the expanding meaning space, as described above, is a much more likely candidate for the growth in complexity seen in the individual condition.

Further to the complexity measures, we also measured the level of iconicity. In previous literature iconicity is generally lost along with complexity as signals conventionalise (Garrod et al., 2007). However, complexity (especially as we are measuring it in this paper) can also arise as the result of combinatorial structure, which has been in inverse relationship with iconicity in some experimental studies (Roberts & Galantucci, 2012; Roberts et al., 2015; Verhoef et al., 2015). We test whether complexity we see growing in our signals throughout the experiment is the result of a reduction in iconicity, hinting at perhaps something like combinatorial structure emerging as a result of the expanding meaning space. However, if iconicity increases it may because of communication driving signals to be more iconic which is aided by complexity in the signals.

We measured iconicity in the same way as Roberts and Galantucci (2012) and Garrod et al. (2007), by getting naïve listeners to match signals with their intended meanings. We found that at the beginning of the experiment, signals in both conditions started with the similar levels of iconicity, though the individual condition was slightly higher. This goes against our hypothesis that the knowledge of another participant would drive signals to be initially more iconic in the communication condition. However, what we found was that signals became more iconic as the communication task progressed. Importantly, this is the opposite of the result of Garrod et al. (2007), where naïve listeners who only saw drawings from the end of the experiment were worse at matching them to their correct referents than naïve listeners seeing the earlier drawings. Again, we can account for this result because of the fundamental differences between our paradigm and the drawings used by Garrod et al. (2007). It is much easier to be iconic with the more flexible drawing paradigm, especially for visual stimuli, allowing for more iconicity at the beginning, which can then be “lost". However, this does not account for the backwards trend we find in our communication condition. It is possible that this result is because of participants becoming more accustomed to the communication game and good strategies to use. Having another participant present with whom you are communicating may be driving the signals to be more iconic. Perhaps, the communication process causes signals to adapt to be more mutually intelligible. While signals produced by an individual for themselves may have a certain level of iconicity (at the levels found in the individual experiment), it is not necessarily true that this iconicity is transparent for naïve listeners. What makes a signal fit for communication may be iconicity that is less idiosyncratic. It may be that signals need to reach this level of transparent iconicity before they can be emancipated from their meanings in order to partake in the process of conventionalisation.

In Perlman, Dale, and Lupyan (2015), non-linguistic vocalisations also became more iconic over the course of a communication game possibly for similar reasons. Both vocalisations, and the signals produced using the leap motion, present a difficulty for producing transparently iconic signals. This difficulty is not so present when using gesture or drawing as modalities that negate the need for an initial stage of negotiation. This explanation makes sense in the light of the signals not gaining iconicity in the individual experiment (see figure 8) because signals can remain idiosyncratic to one person in that condition. Iconicity generally requires more complexity which would explain why signals become more complex.

We also found that participants were much more able to recognise signals within the experiment in the individual condition than in the communication condition. In the individual condition, no negotiation is needed to establish signals, which inevitably leads to higher scores. We also found that in the individual condition, participants got slightly better throughout the experiment, despite the meaning space growing. This could be because participants are simply becoming more used to the apparatus and task throughout the experiment. In the communication condition, participants got worse, probably because the meaning space was growing, making the task more difficult, though as it only expanded by 2 meanings at a time, the effect of having new meanings to negotiate should not have affected the success rate throughout the experiment. However, new meanings competing iconically with old meanings could have affected success for both, and participants did self-report finding some meanings difficult to differentiate (e.g. the spiky brown shape and the white star). Previous artificial language experiments have demonstrated context effects on structure that comes out in these experiments (Winters, Kirby, & Smith, 2015). That is, signals only encode information that is relevant to successful communication which may be different features depending on what other meanings are present. For example, if randomly selected meanings in the recognition task all had shared features this may produce different behaviour and cause specific features to be encoded in signals which wouldn't happen if all meanings had different features. As the meaning space in the experiments presented here are designed to be unstructured and not have any shared features, the effects of context are likely to be much less severe than experiments with structured meaning spaces.

We did not have a condition in our experiment for concurrent feedback, where participants could interrupt one another to initiate repairs, because feedback only came after signals had been completed, transmitted and recognised. Previously, Healey et al. (2007) found that concurrent feedback in a task can be the driving force which makes representations more abstract and less iconic. Garrod et al. (2007) also ran a condition with concurrent feedback, and found that the loss of complexity proceeds faster with ongoing interaction throughout the production of drawings. Participants interrupting each other was also one of the driving forces for conventionalisation in Caldwell and Smith (2012). A paradigm using concurrent feedback may be a worthwhile experiment to conduct using our paradigm. However, as signals are already so short (around 3 seconds), it may not provide much opportunity for interruption, and may in fact drive signals to be longer and more complex so that hearers can be more sure of their guess before interrupting.

Conclusion

We have shown that conventionalisation, as a process for arbitrary forms to emerge, may not work in the same way or as quickly with different modalities. We found no evidence that signals in our experiment became more conventionalised (simpler and less iconic) through interaction or repetition. We hypothesise that when iconicity is difficult in a modality, iconicity needs to emerge over a period of negotiation to gain transparent, mutually intelligible signals. It is only when a signal is grounded for more than one person that it can then be separated in form from its meaning and become more arbitrary. Further, the pressure for discrimination with more restrictive signal spaces may also act against the conventionalisation process causing signals to become more complex. It is not possible with the current work to say which of the above best accounts for our results, but we believe this work is a good first step to demonstrate how modality might affect the process of conventionalisation.

In this article we have compared our results to those of Garrod et al. (2007). However, the current study differed in more ways than only the modality. The expanding meaning space was a confound in our experiment as well as only having visual meanings that will be easier to communicate using a visual modality. An important next step, then, should be to have an experiment with a direct comparison between two conditions where modalities differ only in their flexibility and iconicity.

Acknowledgements

Thanks to Catriona Silvey for thorough comments on a draft of this paper and 3 anonymous reviewers and the editors of this special issue. We would like to thank the financial support of the European Research Council project, ABACUS (283435).

References

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PloS One*, *7*(8), e43807.

Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2016). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science*.

Ehret, K., & Szmrecsanyi, B. (2011). An information-theoretic approach to assess linguistic complexity. *Complexity and isolation. Berlin: de Gruyter*.

Eryılmaz, K., & Little, H. (2016). Using leap motion to investigate the emergence of structure in speech and language. *Behavioral Research Methods*, doi:10.3758/s13428-016-0818-x.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in Human Neuroscience*, *5*, 11.

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.

Goldin-Meadow, S., & McNeill, D. (1999). The role of gesture and mimetic representation in making language the province of speech. In M. C. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution.* (pp. xii, 361). New York, NY, US: Oxford University Press.

Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, *31*(2), 285–309.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 88-111.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1651), 20130298.

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.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Motamedi, Y., Schouwstra, M., Smith, K., & Kirby, S. (2016). Linguistic structure emerges in the cultural evolution of artificial sign languages. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference (evolang11)* (p. 493-495).

Namboodiripad, S., Lenzen, D., Lepic, R., & Verhoef, T. (2016). Measuring conventionalization in the manual modality. *Journal of Language Evolution*, lzw005.

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, *2*(8), 150152.

Roberts, G., & Galantucci, B. (2012). The emergence of duality of patterning: Insights from the laboratory. *Language and Cognition*, *4*(4), 297-–318.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66.

Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural Language & Linguistic Theory*, *29*(2), 503-543.

Schlehofer, D. (2016). *Sign language archaeology: Understanding the historical roots of american sign language*. Oxford University Press.

Theisen, C. A., Oberlander, J., & Kirby, S. (2010). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*, *11*(1), 14–32.

Verhoef, T., Kirby, S., & Boer, B. (2015). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science*.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(03), 415–449.

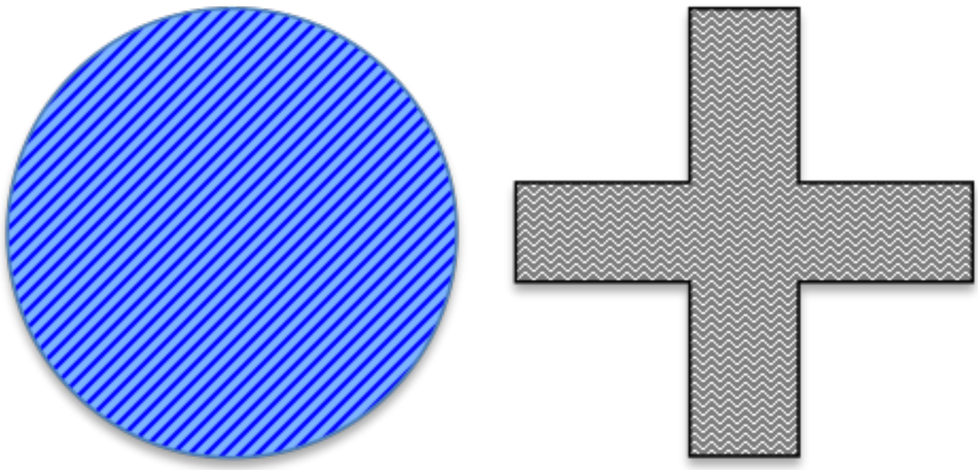

Figure 1. Two meanings with different shapes, colours and textures.

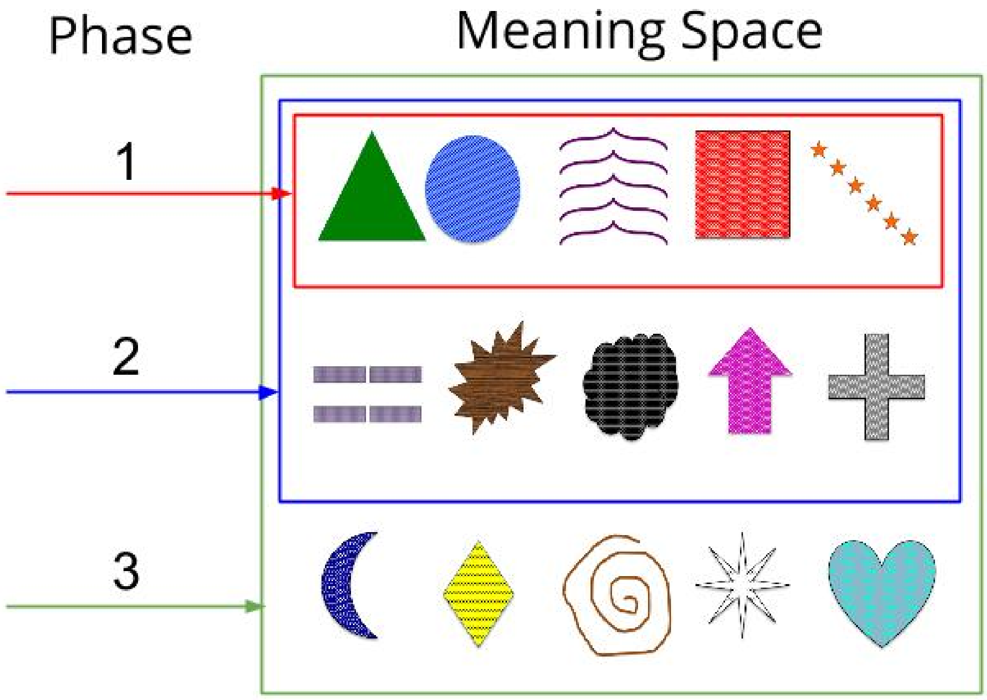

Figure 2. The shapes used as meanings in the experiment in the 3 phases in the individual condition, with the meaning space increasing by 5 with each phase.

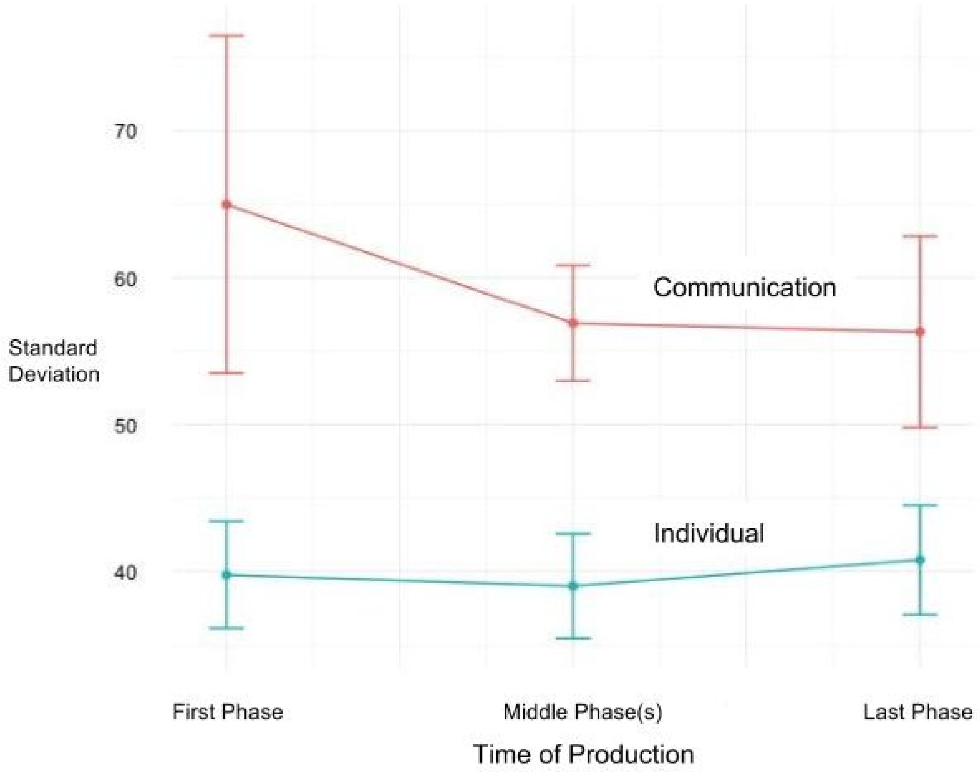

Figure 3. The standard deviation of coordinates within signals, indicating the amount of movement in signals, produced in both conditions, at the beginning, middle and end of the experiment. Here, "last phase" means phase 3 in the individual condition, or the last phase which participants got to in the communication game.

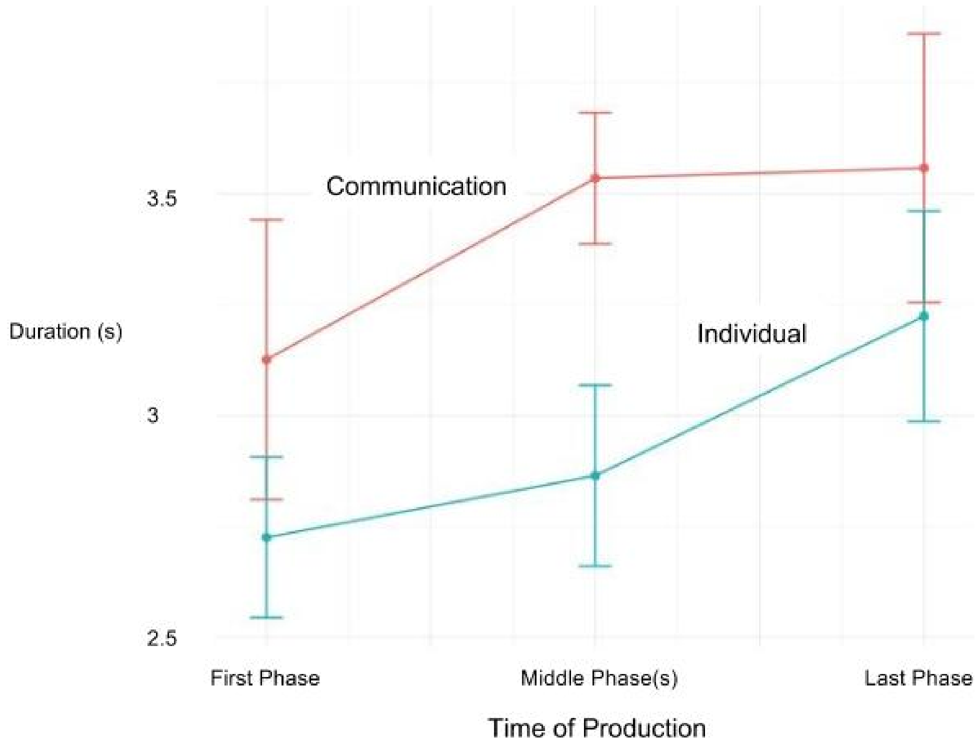

Figure 4. The durations of signals produced in both conditions, at the beginning, middle and end of the experiment. Here, "last phase" means phase 3 in the individual condition, or the last phase which participants got to in the communication game.

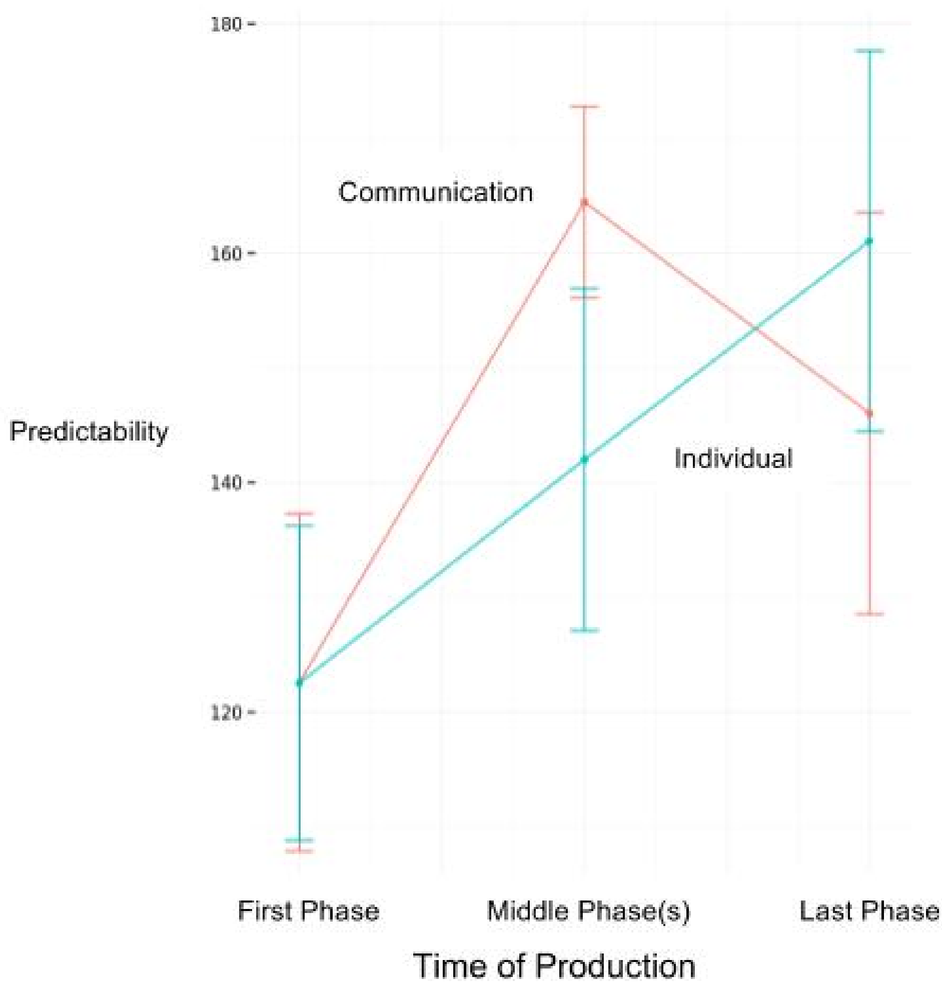

Figure 5. The predictability of signals produced in both conditions, at the beginning, middle and end of the experiment. Here, "last phase" means phase 3 in the individual condition, or the last phase which participants got to in the communication game. Higher numbers here refer to lower predictability (or high complexity).

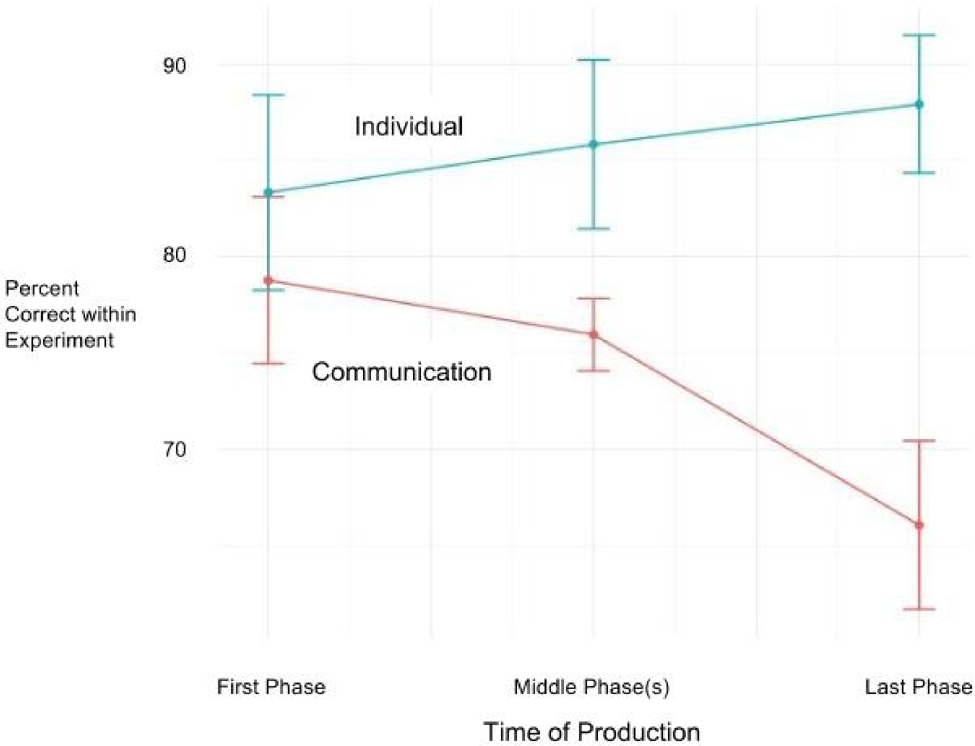

Figure 6. Scores of participants within the experiment, at the beginning, middle and end of the experiment. They are not cumulative, but a sample of responses from phases at the different periods. Again, here, "last phase" means phase 3 in the individual condition, or the last phase which participants got to in the communication game.

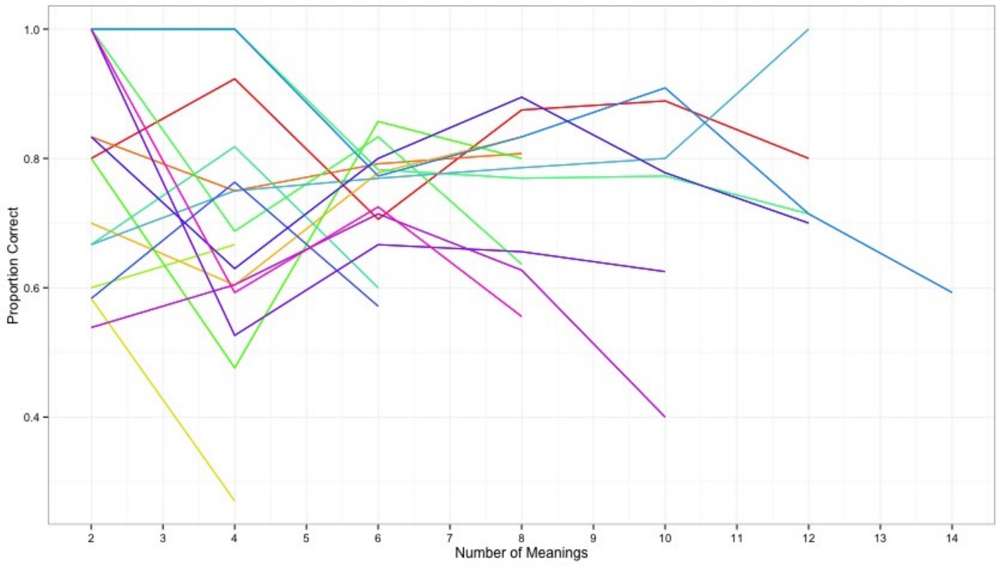

Figure 7. The success of participants throughout the experiment in the communication condition. The scores are not cumulative, but the percent of correct responses within each phase of the experiment, defined by the period before each meaning space expansion. Each pair is one line, and the length of the line illustrates how far that pair got within the 50 minute time limit.

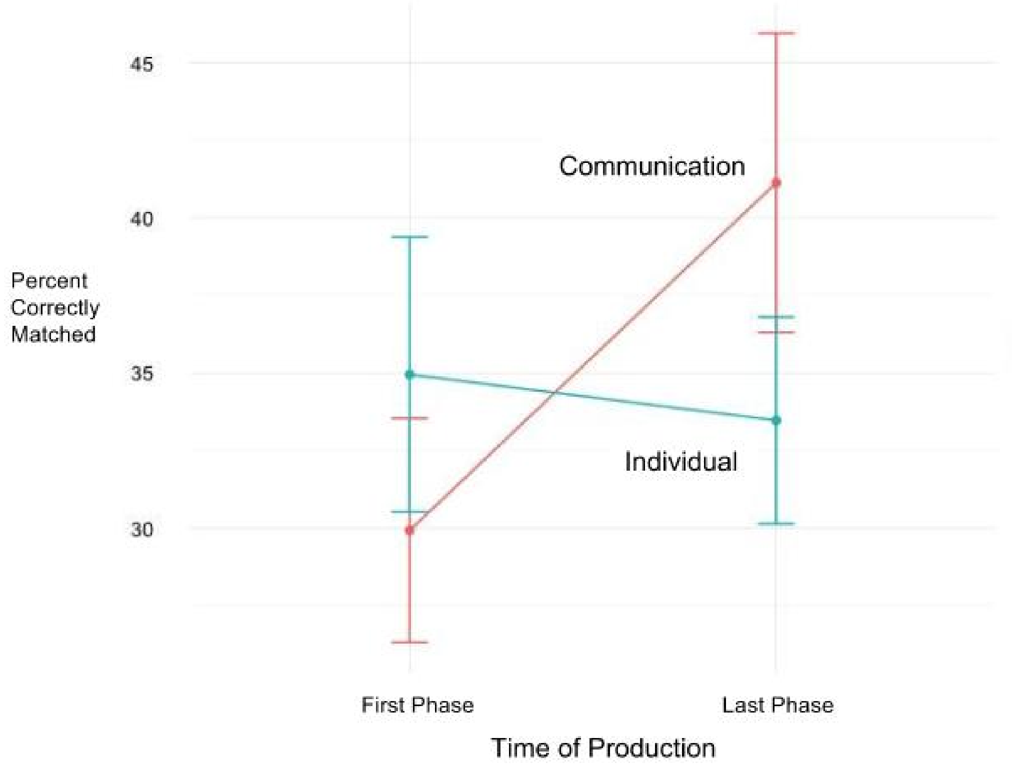

Figure 8. The percentage of signals correctly matched with their meanings by naïve listeners. Both signals produced at the beginning and at the end of the experiment were tested. Here, "last phase" means phase 3 in the individual condition, or the last phase which participants got to in the communication game.

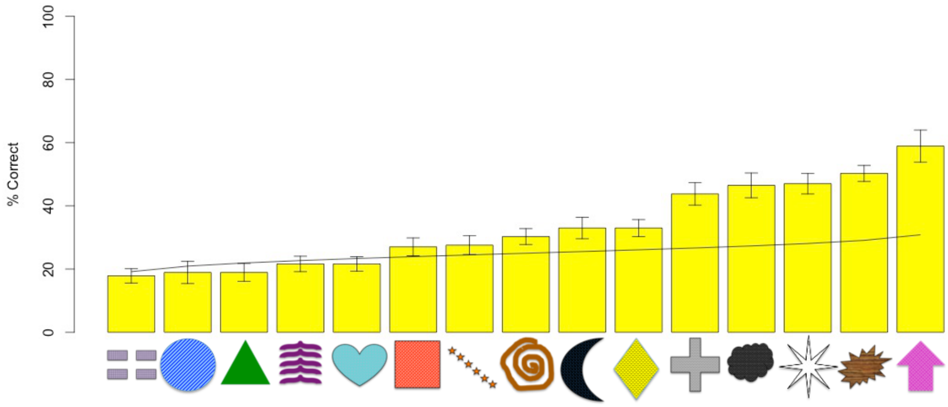

Figure 9. The percentage of correct responses from naïve listeners matching signals with their intended meanings. The graph shows data from the last phase of the individual condition with 90% error bars. The line represents what we would expect if matchers were behaving at chance level.