ANYONE FOR CARDS? EXPLORING THE EMERGENCE OF LINGUISTIC STRUCTURE IN A NOVEL COMMUNICATION SYSTEM USING GAMIFICATION.

Sophie Woodman
Linguistics, 2024

Abstract: The present study ventured to explore the emergence of linguistic structure in the languages created by participants communicating using a completely novel language system, called Ferro, which is completely removed from any pre-existing language. This was achieved through the implementation of a gamified card game wherein participants learned Ferro 'on the fly' as they played in one of two conditions, distinguished by difficulty. Ferros were used to communicate about a defined meaning space made up of Organelles. This card game was used to measure participant accuracy, the emergence of compositional structure, and colour and shape saliency with the aim of answering how successfully, and to what extent, participants created a communication system between them. Results revealed a multitude of findings. Firstly, on average, pairs in the 'hard' condition became less accurate over the course of gameplay such that pairs in the 'easy' condition were considerably more accurate. However, whether increased accuracy was a result of more consistent form to meaning mappings requires closer analysis. Furthermore, there was no emergence of compositional structure in any pair in the 'hard' condition, seemingly due to the lack of compressibility pressure throughout. Instead, all pairs in the 'hard' condition and all but one pair in the 'easy' condition adopted a holistic structure characterised by one-to-one mappings. Finally, shape was much more salient than colour in both experimental conditions. The cause of this can likely be attributed to the organic shape of Organelles visually overpowering the simple colours used in this study.

Keywords: linguistic structure, novel communication system, gamification, Ferro, compositionality, saliency

Supervisor(s): Dr. Christine Cuskley

1. Introduction

Exploring the origins and emergence of linguistic structure is of paramount interest in modern linguistic research. Linguistic structure refers to the systematic arrangement of components of a language which enables clear communication among language users. For many years, research on natural languages has been limited in its ability to answer the question: How does structure emerge in a new language? This limitation is due to the inability for research on natural languages to go back to the beginning of its' use, "as these languages were first formed in the underdocumented recesses of history" (Brentari and Goldin-Meadow 2017: 364). However, in the late 20th and early 21st centuries, new theoretical frameworks have been developed which implement the use of novel artificial languages, many of which utilise said frameworks within the scope of simple sender-receiver settings (e.g. Kirby et al 2015). Such frameworks include the Iterated Learning Model and Closed-Group Settings. The use of novel communication systems within these frameworks provide the tools to bypass this problem, allowing researchers to track the emergence of linguistic structure in a language completely unfamiliar to those using it.

Whilst implementing novel communication systems into investigations of emergent linguistic structure provides a solid and rich basis upon which to investigate participants' linguistic behaviour, such methods are not free from bias. Previous research from Kirby et al (2008; 2015) and Raviv et al (2019) explore the emergence of compositional structure using novel communication systems. However, the novel labels that constitute the form spaces in these studies are largely formed using a CVCV structure and are comprised of letters drawn from the English alphabet. To English speaking participants, these structures and letters will be inherently familiar. In an attempt to mitigate this bias, this study utilises a completely novel language system, called Ferro, which is totally removed from participants' prior knowledge of a language. In doing so, this study aims to answer the question: When presented with a completely novel language system, how effectively will participants create a communication system between them? This is achieved through the implementation of Ferro, which is learned by participants over the course of a card game. Through this learning process, the emergence of structure in participants' novel languages is investigated.

Particular emphasis was placed on whether contrasting levels of game condition difficulty would result in the emergence of compositional structure in participants' communication. Of the two conditions, one was considerably harder than the other.

The hypotheses of this study were the following:

1. Accuracy of participants' turns will increase over the duration of gameplay, in both experimental conditions, as a result of consistent form to meaning mappings between Ferros and Organelles.

2. Compositional structure will emerge in the languages of participants playing in the 'hard' condition given the increased difficulty distinguishing between Ferros leading to a constraint on memory.

3. The colour of Organelles will be more salient than the shape and, where compositional structure emerges, reference to colour and shape will follow the perceptual saliency hierarchy (PSH).

The results relating to the first hypothesis partially falsify it. Whilst accuracy increased in the 'easy' condition throughout gameplay, in the 'hard' condition, pairs became less accurate over time. Furthermore, in reference to the second hypothesis, compositional structure did not emerge in any of the 'hard' pairs thus finding this hypothesis to be false. Finally, the results proved the third hypothesis untrue based on findings which showed shape to be considerably more salient than colour in both conditions.

The following section begins with a targeted review of literature that considers gamification, and its' effects thereof, in the scope of experimental work, as well as previous literature that has explored the emergence of linguistic structure in novel communication systems. It is from this literature that the hypotheses of this study were born. Subsequent sections address the experimental methods utilized in this study before presenting the results and a discussion of this paper's findings.

2. Background and Literature Review

2.1 Gamification

The term gamification has gained considerable traction since it first emerged in the early 2000s (Sailer et al 2017). At its centre, gamification refers to the implementation of game-like elements in non-game contexts (Deterding et al 2011). Sailer et al offer a simple yet encompassing definition of gamification, where they suggest that the “central idea is to take the ‘building blocks’ of games and implement these in real-world situations” (2017: 371)

Though gamification as a formalized concept is a relatively new phenomenon, traces of gamification can be found as far back as 1896, with S&H grocery stores in America implementing a green-stamp loyalty scheme for customers (see Christians 2018). Christians’ (2018) comprehensive review of the history of gamification shows that such ideas have endured throughout time, spanning from the Boy Scouts badge system in the 1970s to Nintendo’s video game entertainment systems in the 1990s. At the dawn of the digital age, video games became the widely agreed entertainment system of the future, thus transforming our definition of games (Sailer et al 2017). In the 21st century, gamification has been used to implement modern game elements, such as points, leaderboards and badges (Rapp et al 2018), into non-game settings including education (Caponetto et al 2014), marketing (Huotari and Hamari 2016), crowdsourcing (Morschheuser at al 2016) and academia (Muthiyan et al 2023), to name a few. This practice often evokes specific psychological effects in participants such as motivation, teamwork, enjoyment, and flow (Koivisto 2017; Huotari and Hamari 2016). Undoubtedly, as the definition of what constitutes a ‘game’ changes, and impressions about what makes these modern games so entertaining strengthens, a window of opportunity opens in which individual game mechanics can be specifically levied for use in non-game contexts. Ultimately, this practice impacts upon individual motivational outcomes (Krath et al 2021), which can be favourable in experimental work.

Recent scholarship investigating the effectiveness of gamification has reported a mix of positive and negative results. Muthiyan et al 2023 utilised an innovative card game among undergraduate medical students with the aim of aiding the comprehension and memorisation of complex anatomical knowledge. They found that the implementation of the card game was very effective and, crucially, more effective than small group discussion. Students who used the card game saw their test scores increasing from pretest to post-test, demonstrating their increased comprehension and memorisation of the topic (Muthiyan et al 2023). Muthiyan et al's findings converge with those of other empirical studies of gamification in education, which generally report positive outcomes (Hamari et al 2014). Contrastingly, positive effects of gamification have not been universally reported across different sectors. Downes-Le Guin et al (2012) measured respondent engagement with a range of online surveys. Participants were fielded in 4 different survey styles, one of which was gamified. Revealingly, they found no significant difference in respondent engagement measures in the gamified style, equally finding no difference in response patterns (Downes-Le Guin et al 2012). However, respondents who were assigned the gamified surveys reported higher satisfaction scores, that is to say, they experienced greater enjoyment than respondents completing other styles of survey.

Sailer et al (2017) address the conflicting nature of gamification research, reporting that previous studies oftentimes fail to acknowledge that the implementation of different design elements inevitably produce varied psychological effects. In their own study, Sailer et al posit that gamification as a broad tool is not effective per se, but that the specific psychological effects resulting from the implementation of specific game design elements are effective (2017: 371). In sum, different game design elements can be harnessed and used to evoke different psychological outcomes, ultimately influencing a participant's performance on a given task. These psychological effects can positively influence participant experience, or negatively affect it (Hamari et al 2014). Literature appears to converge on what the most important psychological effects in gamified environments are: motivation, mastery, feelings of autonomy, enjoyment, competition and cooperation (McGonigal 2012; Lee 2011; Koivisto 2012; Sailer et al 2017)

Motivation is a crucial effect of gamification that is particularly powerful in encouraging participants' engagement with an experimental task or activity (Sailer et al 2017). Specific game design elements such as points, levels, badges and other physical rewards are particularly successful in initiating, and ensuring the continuation of, goal-directed behaviour through targeted and, in the case of points, instantaneous feedback. These game design elements are easily included or extracted from the gamified environment, allowing for manipulation of the gamified setting and specific effects to be measured. Effects range from the motivation of participant performance, defined as the desire to supersede the standards set by peers, and participants' personal mastery of a task, which involves setting and striving towards self-defined standards (Sailer et al 2013). Motivation is at the root of success in gamified environments. Ryan and Deci (2000) first proposed splitting types of motivation into two distinct categories, intrinsic and extrinsic motivation, as part of their Self-Determination Theory (SDT). Intrinsic motivation arises from participants' personal enjoyment of a task, one which promotes their personal autonomy (feelings of self-directed control over their actions), mastery, and purpose (Ryan and Deci 2000). Extrinsic motivation relates to specific outcomes, such as rewards, that behave as powerful motivators for participants to take part in a task. Recent literature has demonstrated the power of gamification in promoting intrinsic motivation (Rantinho and Martins 2023). Gamified environments promote enjoyment through the satisfaction of participants' needs to feel autonomous, masterful and purposeful. Coupled with the addition of game elements that instil extrinsic motivation, gamification provides a rich learning environment for participants.

Feelings of competition or cooperation can be fostered by the implementation of specific game design elements, such as leaderboards or the formation of teams that work together towards a common goal, within the gamified environment. Werbach and Hunter (2012) classify leaderboards as effective motivators, allowing participants to relate their own performance to the performance of others. However, the overarching motivational potential of leaderboards has been scrutinized, with some scholars reporting them to be equally demotivating if participants find themselves at the bottom of a leaderboard (Sailer et al 2017). The introduction of defined groups of participants that work towards shared objectives (teams) has been reported to consistently and effectively induce feelings of cooperation between participants (Sailer at al 2017). Gamified environments that incorporate these design elements have a direct impact on learning outcomes by fostering cooperation between participants and promoting problem solving, which are often crucial to task success (Thuairasu 2022).

The effects of gamification mentioned above have the crucial consequence of reducing demand characteristics throughout the course of an experiment. The construct of demand characteristics refers to a participant's awareness of the investigatory aims of the experiment they are taking part in (McCambridge et al 2012). Of course, the immediate problem this presents is the potential for participants to alter their behaviour due to an awareness that they are being 'studied'. The consequence of this, from an experimenter's perspective, is that we cannot be certain that we are finding out participants' natural reactions to specific stimuli (Orne 1996). In mitigating these impactful consequences, gamification is a powerful tool. As discussed in length above, the incorporation of game design elements into an experiment endows participants with feelings of motivation, autonomy, and enjoyment which ultimately reduces attrition by bolstering participants' engagement with the game at hand (Huber et al 2023). Such engagement and focus on the gamified environment subsequently distracts participants from attempting to uncover the investigatory intention of the study. McCambridge et al suggest that, in distracting participants from the true nature of a study, we can "prevent demand characteristics introducing unwanted influences on responses" (2012: 2). Nevertheless, it is crucial to avoid lowering demand characteristics to such an extent that participants diverge from an experiment's goals such that their data is useless in the scope of the experiment. Hence, the incorporation of specific design elements that keep participants in line with a study's aims is crucial to gamified experimental work in this field.

The present study utilises gamification to investigate emergent linguistic structure in a novel communication system using a card game. This medium allows for the purposeful incorporation of game design elements to promote intrinsic and extrinsic motivation, ultimately promoting learning. The nature of the card game allows participants to make autonomous and unabated decisions about their method of gameplay. As such, they were able to incrementally progress and improve throughout the game as they headed towards mastery. To complement the natural intrinsic motivation present in the gamified environment, the addition of extrinsic motivators, such as a points-based game element, a leaderboard, and rewards encouraged specific goal-directed behaviour and ensured continuous motivation throughout. The implementation of teams had a similar effect by fostering a cooperation between pairs of participants to compete against other pairs. All of these factors reduced demand characteristics in this study, which resulted in the observation of more naturalistic behaviour from participants.

2.2 The Emergence of Linguistic Structure in Novel Communication Systems

The motivation for this study was born from debates in previous research investigating the emergence of linguistic structure in novel communication systems. Laboratory studies exploring the emergence of structure in human behaviour can be dated back as far as 1932, when Bartlett first utilised his 'serial reproduction' paradigm, which has since been developed into the Iterated Learning Model widely used in modern research (Kirby et al 2014). In recent years, studies have sought to rigorously explore the effects of learning pressures such as compressibility and expressivity on the structure of emergent novel languages (Kirby et al 2015; Raviv 2019). While these experiments have been traditionally carried out with human participants, an increasing volume of work taking a deep learning approach has emerged. These studies have readily implemented the use of neural network agents and Bayesian agents, as well as humans, to further our understanding of emergent linguistic structure (Raviv et al 2019).

Regardless of the nature of participants (whether they are human or robotic), the methodological framework adopted by experimenters remains largely consistent (Galke and Raviv 2024). Experiments typically take the form of sender-receiver games which allow for repeated rounds of communication from which the emergence of linguistic structure has been tested and observed (Galke and Raviv 2024). These sender-receiver games are most often implemented into one of two experimental methods, the Iterated Learning Model or closed-group settings. Iterated learning is the "process by which the output of one individual's learning becomes the input to other individuals' learning" (Smith et al 2003: 371). Often this transmission occurs across many generations of participants, resulting in "miniature trajectories of language evolution" (Cuskley 2019: 2). The well-known poverty of the stimulus is a natural effect of iterated learning models, which encourages compositional structure. This structure is defined by the principle that the meanings of complex expressions are defined by the meaning of their constituent parts (Barrett et al 2020: 911). It is widely regarded as a requirement of natural languages, as it allows the signaller to express a limitless number of complex expressions (Saldana et al 2019). Conversely, close-group settings refer to communication without transmission. In these closed experimental communities, communication occurs exclusively in a single generation (Raviv et al 2019).

There exists some manner of debate across the literature regarding whether compositional structure can emerge in closed-group settings, or whether the emergence of this linguistic structure is only possible when generational transmission is implemented. The key aspects of this debate are explored below.

In an influential study exploring the emergence of linguistic structure, Kirby et al (2015) implemented both the iterated learning model and a closed-group setting into a comparative study. Participants were asked to communicate using an artificial language where words were formed using CVCV strings, akin to English. In the iterated learning model, the language outcomes of the original participants were transmitted to a new set of naïve participants, thus forming a generational chain of transmission. On the other hand, in the closed-group setting, there were no new participants introduced. Therefore, the same participants played for continuous rounds, with their language confined to a single generation. Results from the study found that compositional structure emerged over the course of multiple generations in the language of participants in the iterated learning model. Meanwhile, in the closed-group setting, the language of participants did not become compositional, instead remaining largely holistic throughout the course of the experiment. Kirby et al (2015) argue that compositional structure emerged in the language of participants communicating with generational transmission due to a trade-off between compressibility pressures (making language more simple and, therefore, more learnable) and expressivity pressures (ability for straightforward one-to-one mapping between forms and meanings). Compositional structure emerges here as a simple solution to both pressures being exerted on the language (Kirby et al 2015: 98). The authors go on to comment that the introduction of fresh, naïve participants in each round of the chain transmission prevented participants from simply relearning in a previously established style, as in the closed-group setting. Kirby et al (2015) suggest that this pattern of relearning occurred due to the lack of compressibility pressure in the closed-group setting, resulting in the continuation of a holistic language structure. Based on these findings, Kirby et al (2015) claim that compositional structure emerges as a trade-off between compressibility and expressivity pressures, and that these are both exerted only in communication with generational transmission.

Despite the results of Kirby et al's (2015) study, other studies in the field have presented opposing findings, which indicate that compositional structure can emerge in single generation communication settings (Selten and Warglien 2007; Raviv 2019). Selten and Warglien (2007) conducted an experiment that focused on dyadic communication between pairs of participants. Participants were required to assign a string of acceptable consonants to a geometric figure that randomly appeared on a computer screen. Geometric figures were presented in sets and successful transmission was defined by matching codes. Results showed that 12% of participants formed compositional codes where strings of letters were combined to reflect shape, colour, and insert (see Selten and Warglien 2007). Through this experiment, Selten and Warglien (2007) show that compositional structure can emerge not only in a single generation, but also in dyadic communication. Raviv et al (2019) highlight this further in a more recent study, where they created an environment in which two pairs of participants communicated with each other (four-way communication) in an artificial language. Participants used an artificial language to refer to novel scenes. These scenes constituted the meaning space which, crucially, was expanding throughout the experiment. Raviv et al (2019) explored whether these two compressibility pressures would adequately trigger the emergence of compositional structure in a single generation. The results showed that the languages formed by the closed-groups of participants became increasingly structured over time and compositional structure emerged despite the lack of generational turnover (Raviv et al 2019: 159). These studies challenge the conclusions drawn by Kirby et al (2015) in that they demonstrate that with sufficiently implemented compressibility pressures, compositional structure can emerge in novel communication systems despite a chain of transmission not being utilised. One shared characteristic of the above studies is that the languages formed by participants became more accurate, converged upon, and structured over time (even if minimally in the case of Selten and Warglien 2007), which logically presupposed the emergence of complex linguistic structure. This is the line of thinking behind the hypothesised increase in accuracy over time due to consistent mapping between form and meaning in this study.

One unanimous feature of previous studies, regardless of whether they implement communication with or without transmission, is that the novel language utilised in the studies are largely signals created using written words or letters, such as in the literature above. More often than not, these studies are conducted with English speaking participants, resulting in experimental stimuli being produced to resemble words that could plausibly occur in English, or are indeed English pseudowords. The aim of the current study is to explore the emergence of linguistic structure in a novel communication system that is removed from artificial words or pseudowords formed using letters. Instead the novel communication system in this study takes on the form of Ferros (see the Materials section) which bear no resemblance to letters or words, thus minimising any attempt at bootstrapping known language systems. The two experimental conditions in this study, 'hard' and 'easy', have expectedly different effects on participants' memory loads. For those in the 'hard' condition, Ferros were incredibly challenging to distinguish from one another when not placed side by side. Therefore, memory constraints on visual information paired with the "fast and fleeting nature of linguistic input" (Christiansen and Chater 2016: 2) imposes a Now-or-Never bottleneck which Christiansen and Chater (2016) describe as a requirement for input to be processed immediately, else risk being overwritten by new information. In this study, participants are exposed to form-meaning mappings (Ferros to Organelles) for a finite amount of time. There is then an interval between that mapping and their next exposure. During this time, information is inevitably lost. In turn, this bottleneck effect amplifies learners' convergence on compressible (e.g. simple) structures (Kirby et al 2015), which ignite the push towards compositional structure in closed-group settings. Given that this theory applies most strongly to the 'hard' condition, the hypothesis is that compositional structure will emerge in participants' formed languages in the 'hard' condition.

In addition to compressibility and expressivity pressures, research has shown that perceptual saliency could have an effect on the structural configuration of language (Gong et al 2016). Pedale et al define perceptual salience as "the distinct subjective perceptual quality which makes some items more attentional-capturing than others" (2022: 1). Previously, perceptual salience has been found to affect short-term memory "whereby perceptually salient objects have more chances to be encoded and then successfully retrieved than objects with lower salience" (Pedale et al 2022: 1). In further support of this view, Gong et al conducted an artificial language learning experiment which explored whether the perceptual saliency hierarchy (herein PSH), where colour is more salient than shape which is more salient than texture, could influence the learning or processing of language structures (2016: 1). Results showed that participants exhibited biases towards the orders that were congruent with the PSH, confirming that the PSH can affect individual learning of structure. In this study, Organelles (meanings) take on two perceptual characteristics, colour and shape. Based on the literature above, its hypothesized that colour will be more salient than shape. Furthermore, where compositional structure emerges, the hypothesis is that the structure of signals produced will be congruent to the PSH, that is to say, colour will come before shape.

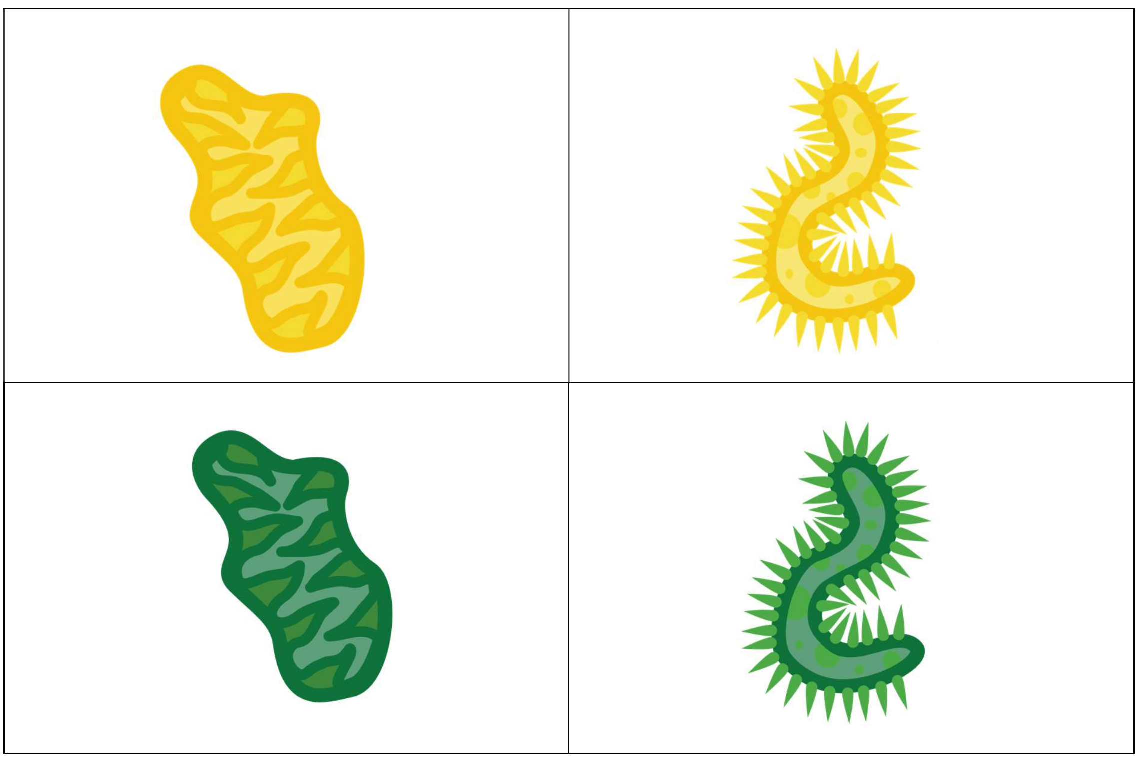
3. Methodology

3.1 Participants

The participants in this study were personal friends who volunteered to take part in this experiment. As such, participants in this study knew each other prior to the commencement of the experiment. In total, there were 18 participants in this study: 9 men and 9 women, with participants' ages ranging from 18 to 21 years old. The median age of the group was 20 and the mean age was 20.2. Most participants were students at either Newcastle or Northumbria Universities. Only 2 participants were in full-time employment. Participants were told that there would be a leaderboard, which ranked pairs based on points scored during gameplay, resulting in two winning teams, one from each condition. This study was considered 'low risk' and informed consent was given by the participants by way of consent forms. Subsequently, a total of 9 recordings were made between 5th January 2024 and 1st February 2024. The information sheet for this study, which was provided to participants before the commencement of gameplay, can be found in Appendix 1.

3.2 Materials

Participants used a printed set of novel playing cards during gameplay[1]. This set consisted of 10 meaning cards called 'Organelles' which each appeared once in the centre of gameplay and a further 10 times in the playing deck. In total, there were 110 Organelles. These varied in colour (blue, red, green, yellow, purple) and shape (spikey, blob). Illustrations of each Organelle can be found in Figure 1. Furthermore, the set included 20 form cards, namely 'Ferros' which were drawn from Cuskley's (2019) study on novel form spaces in cultural evolution (see Cuskley 2019 for information regarding the creation of the Ferros). Ferros were split into two conditions, 'easy' and 'hard', with participants playing with just one set of 10 Ferros, dependent on which difficulty level they played at.



[1] The novel deck of playing cards was printed by the British Academy in conjunction with a project by Cuskley (2019) in collaboration with the British Academy.

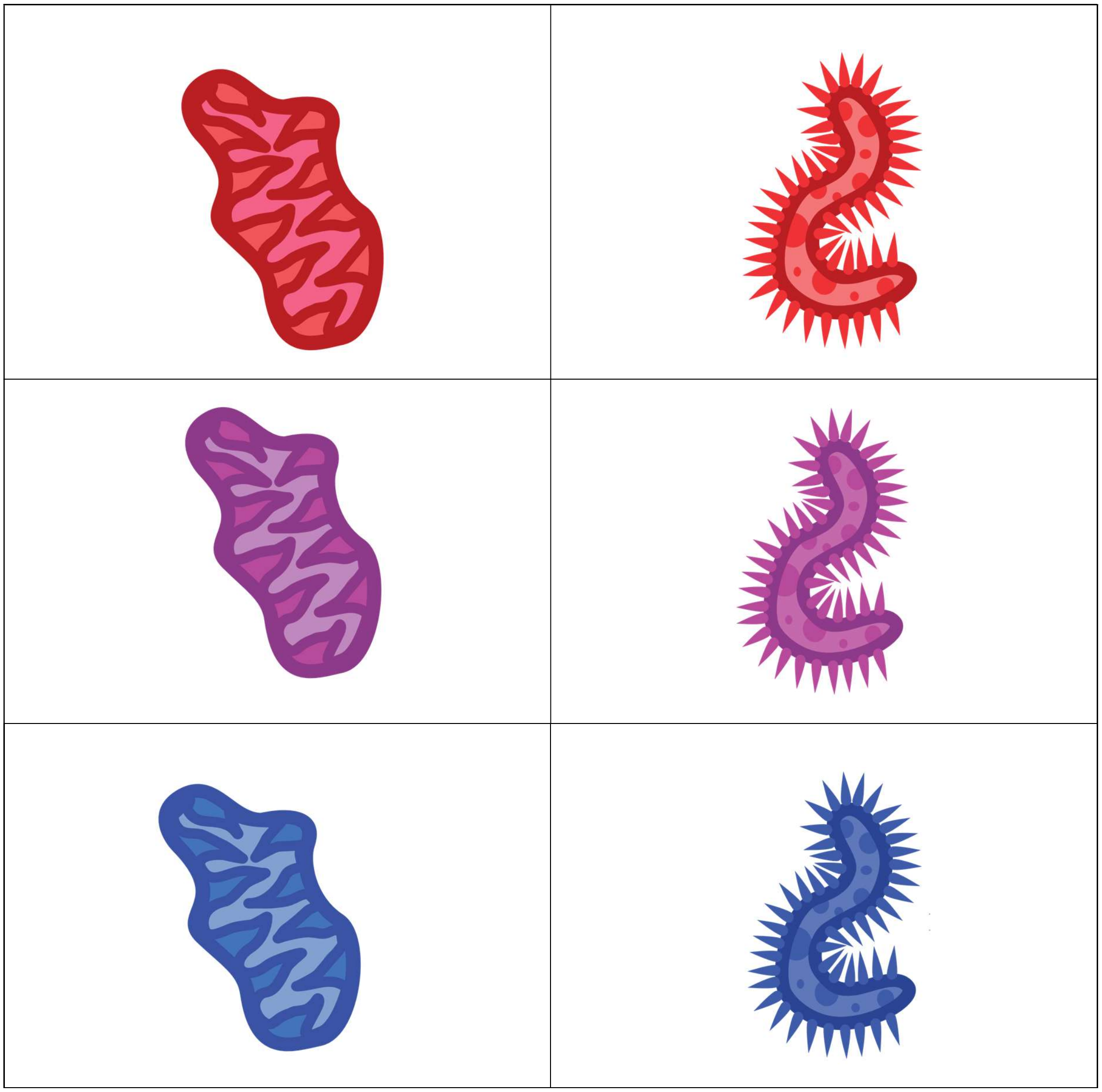
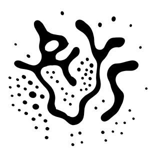
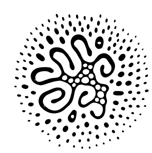
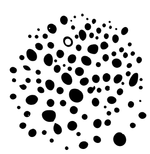
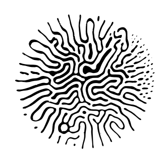
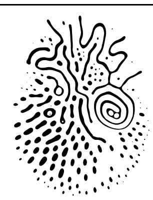
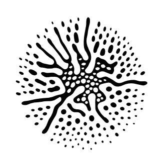
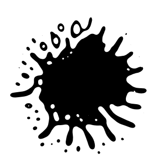
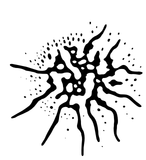
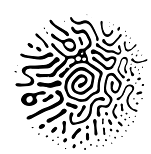


Figure 1: A visual representation of the 10 meaning cards, or Organelles, used in this study

To communicate, participants 'played' (displayed) Ferro cards during their turn, choosing from any of the 10 Ferros in their hand, which differed depending on whether they were playing in the 'easy' condition or the 'hard' condition. As described by Cuskley (2019), Ferro's are unique as a method of communication due to their complete dissimilarity to any other communication system, such as letters. Ferros belonging to the 'easy' and 'hard' conditions differ with respect to the ease with which one can distinguish between them. Ferros were considered 'easy' if they were easily distinguishable from other Ferros, such that participants would not struggle identifying Ferros apart from one another. Though dissimilar, it remained evident that these Ferros were all part of the same set. Meanwhile, Ferros labelled 'hard' were characterised by their similarity. They were not easily differentiated unless they were positioned side by side. Thus, mapping between forms and meaning (Ferros and Organelles) in the 'hard' condition would be more challenging. See Figure 2 for a full table of the 20 Ferro's used in this study.

Easy	Hard
	
	
	
	
	
	

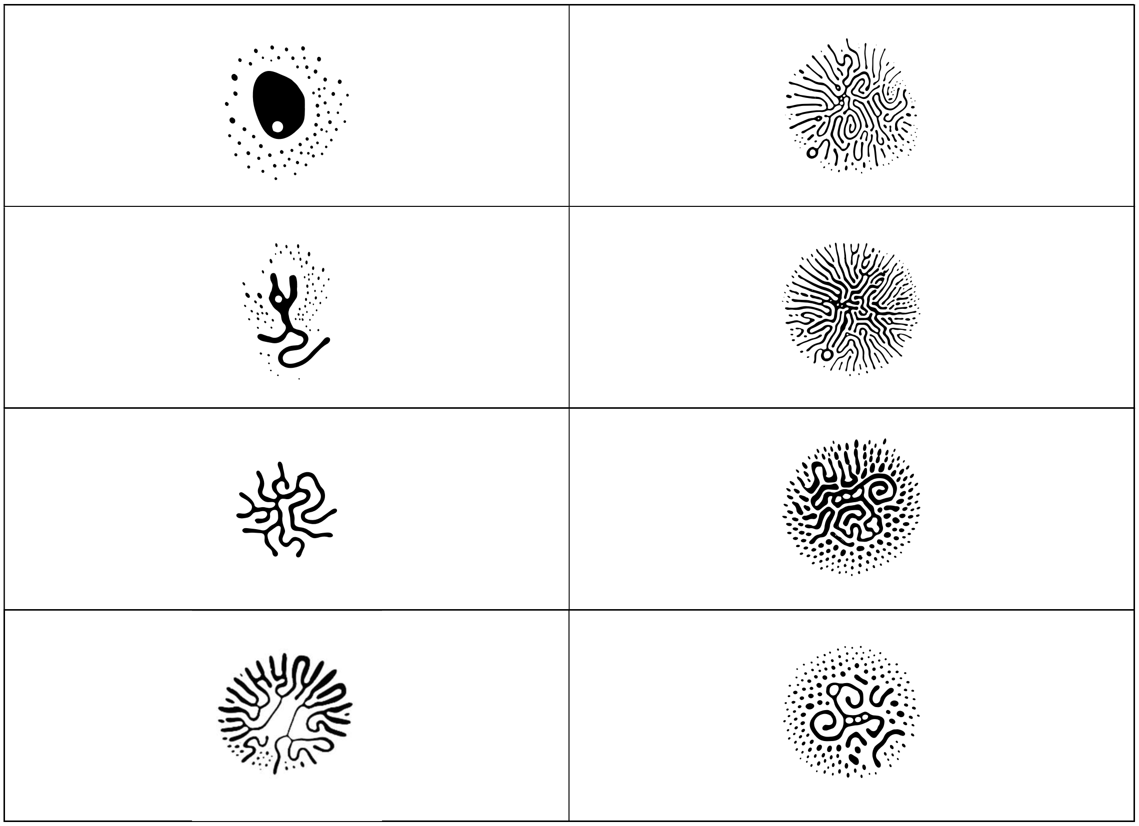


Figure 2: A visual representation of the 20 form cards, or Ferros, used in this study

This experiment took place in the LingLab on Newcastle University's campus. Further equipment used in this experiment consisted of a tripod, from which a long piece of plastic was secured. At the end of this plastic, we attached a GoPro camera with the lens facing downward so as to capture gameplay from above. This ensured that all aspects of participants' gameplay could be recorded, while keeping any identifiable features, such as participants faces, out of frame. A small sheet marked 'SIGNAL' was placed in the centre-bottom of the camera frame to ensure that when participants produced Ferros, they were always visible in frame. Participants were also instructed to bring a pair of noise-cancelling headphones to listen to their own music, which they wore throughout the duration of gameplay. A pair of backup headphones was on hand in the event that a participant forgot to bring their own.

3.3 Procedure

Before the commencement of gameplay, participants read the provided information sheet and consented to participation and having their gameplay recorded. Subsequently, participants were given verbal instruction, where the rules of the game were explained to them. They were informed that they would be playing a communication-based card game together for 30 minutes, and that they should endeavour to score as many points as they could in this time frame, as they would be working together. They were asked not to communicate with each other for the duration of the experiment in any way beyond using the cards. This included non-verbal communication, such as hand gestures or eye movements, as well as verbal communication. In an attempt to reduce the temptation to communicate, participants were asked prior to arriving at the LingLab to bring noise cancelling headphones with them. Participants were permitted to listen to their own music for the duration of the experiment. The whole experiment lasted between 35 and 50 minutes.

Participants sat opposite one another, either side of a table. Between them, all 10 Organelles were placed face-up in a random, circular formation. This prevented participants from drawing any immediate connections between Organelles that shared the same colour or shape or using their eye-gaze to orient the attention of their partner. The remaining 100 Organelles were stacked face-down in a deck to the side of the participants in a random order that was fixed for all games, allowing the experimenter to determine the target meaning in each turn, even where it was out of frame. A table containing the full order of Organelles can be found in Appendix 2. Participants drew Organelles from this deck during gameplay. Each participant was then handed a set of Ferros which belonged to either the 'easy' or the 'hard' difficulty level. Both participants played with an identical set of Ferros, either both playing with 'easy' Ferros, or both playing with 'hard' Ferros. The difficulty level at which pairs were playing was randomised, with each pair being assigned a difficulty level prior to gameplay. A visual image of the experimental set up is provided in Figure 3.

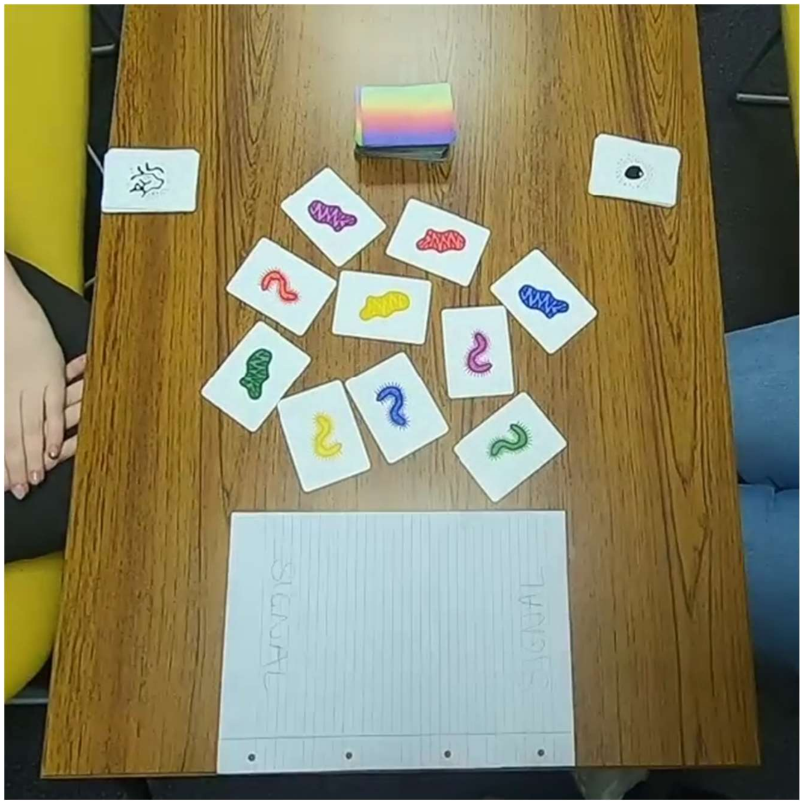


Figure 3: Visual image of the experimental set-up prior to gameplay.

A decision regarding which participant would draw first was spontaneously decided between the participants prior to the first turn, and gameplay was initiated by the first participant drawing their first Organelle from the top of the deck. Therefore, this participant was the first to signal with their Ferros, and thus became Participant 1. Their partner, who will receive Participant 1's signal, became Participant 2. Having drawn their Organelle, Participant 1 would assess the 10 Ferros in their hand to determine which one(s) they believed could appropriately convey to their partner the Organelle they had drawn. There was no time limit on this process. Once decided, Participant 1 was required to play their Ferro(s) onto the 'signal' sheet (see Figure 3). Participant 2 was then able to assess the Ferro(s) Participant 1 had played, subsequently pointing to the Organelle on the table they thought Participant 1 was intending to communicate. If the Organelle selected by Participant 2 was the same as the Organelle drawn from the deck by Participant 1, then the pair scored a point. Participants were instructed to place correctly communicated Organelles face-down in a pile, to the side of the main playing zone. If the selected Organelle was incorrect, Participant 1 would place the drawn Organelle face-down behind the 'live' deck. On the rare occasion Participant 2 wanted to amend their original choice of Organelle, their first choice remained the prevailing one. Participants were informed of this before gameplay and were reminded of this rule, should it have been necessary. Whether correct or incorrect, Participant 1 collected their played Ferro(s) from the 'signal' sheet and returned them to their hand, signalling the end of their turn. At this point, Participant 2 drew a new Organelle from the deck, and the process was repeated. This turn-taking structure continued until the end of gameplay. Once finished, the points accrued by each pair were added up and noted down.

3.4. Analysis

Data was obtained from each video recording by coding using ELAN[2]. This involved individually coding each video recording. In total, each recording was coded for the same 11 tiers. Figure 4 shows the first 15 coded turns from Easy Pair 3's gameplay. Each tier is visible.

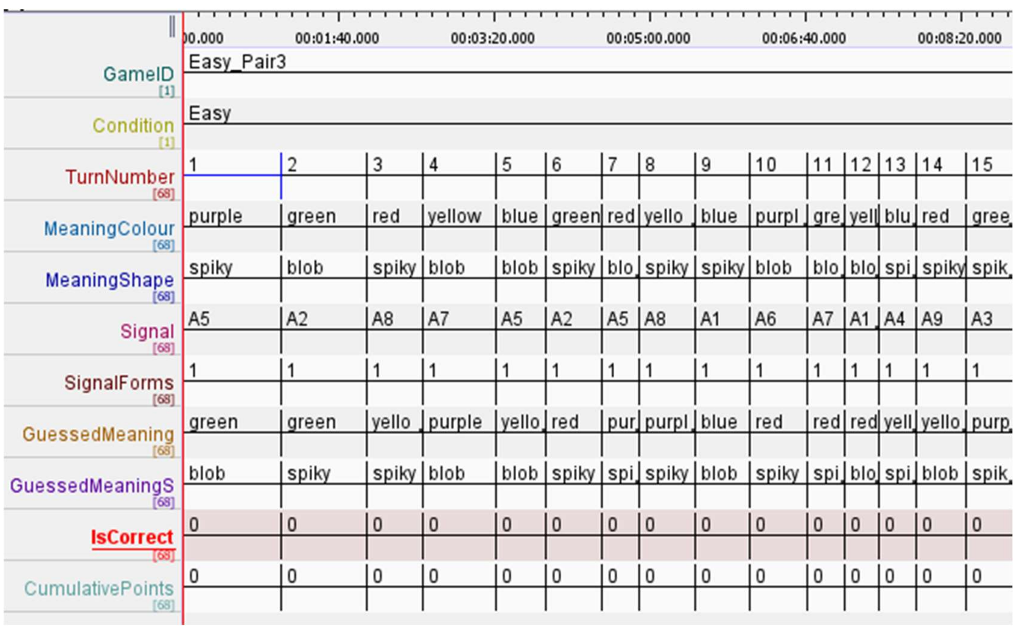


Figure 4: Easy Pair 3's first 15 turns as they were coded in ELAN.

The first tier coded for each video was 'GameID', followed by 'Condition'. These were the only two tiers that spanned the entire length of gameplay. These tiers were coded to begin

[2] Version 6.7 (the most up to date version of the software at time of writing).

when Participant 1 picked up their first Organelle from the deck. It ended when the last participant completed their turn and placed down their Organelle, which was after 30 minutes of continuous gameplay. The next tier to be coded was 'TurnNumber'. This involved highlighting, in ELAN, the time spent on a turn and subsequently labelling that turn '1', '2', '3' and so on. The time a turn starts was characterised by a participant picking up their Organelle from the deck. The end of a turn was coded at the moment the participant placed their Organelle in the relevant pile ('correct pile' or 'incorrect') after their turn was complete. Where the second participant picks up their Organelle before the preceding participant places their Organelle down, the previous turn ended when the second participant picks up their new Organelle and the new turn begins.

'MeaningColour' and 'MeaningShape' were coded to reflect the actual characteristics of the Organelle the participant was intending to communicate. This information was obtained through a table, like in Appendix 2, which showed the pre-set order of the Organelles with which pairs would be playing. 'GuessedMeaningColour' and 'GuessedMeaningShape' reflected which colour and shape participants actually chose, irrespective of whether they were correct. Signal showed which Ferro(s) participants used to communicate with their partner. Their letter and number label combination was used here. These were consistent between each pair. Ferros were ordered the same as they are displayed in Figure 2. Subsequently, those in the 'easy' condition were labeled A1 to A10 and those Ferros in the 'hard' condition were labelled B1 to B10. As such, the first Ferros in Figure 2's respective Easy and Hard columns are A1 and B1. 'SignalForms' showed how many Ferros a participant used to communicate during each turn, whether there was just one used or multiple and if there were multiple, how many were used. 'IsCorrect' and 'CumulativeScore' simply reflected whether pairs guessed correctly and scored a point, and what that running total of correct answers was throughout gameplay.

Once all videos had been coded in ELAN, they were collated together to form one large data frame which was subsequently exported to R studio for analysis. There were 4 pairs in the 'easy' condition and 5 pairs in the 'hard' condition. Upon beginning analysis, it became evident that one pair, Hard Pair 3, were an anomalous result. This was due to their gameplay having been largely influenced by their reliance on established language systems, which will be elaborated upon further in the Results. As a result, Hard Pair 3's data was excluded from analysis from this point onwards. With this pair's data removed from the data frame, there was a total of 563 turns taken across all participants in both conditions.

The results of gameplay were calculated across various measures. The first result given by the experiment was accuracy, which was calculated as a percentage by taking each pair's total number of points scored and dividing it by the total number of turns each pair took. In order to measure the hypothesised increase in accuracy over time, the cumulative frequency of each pair's results, which was calculated by placing a pair's number of points against each completed turn and only increasing this number when a turn was correct, was plotted linearly against the number of turns each pair took. From this, accuracy across conditions could be measured as it was hypothesised that those pairs in the 'easy' condition would be more accurate than those playing in the 'hard' condition, thus scoring more points. While there were 100 points available for each pair in each condition, visualisations of results were cut appropriately to reflect the overall number of points scored in the data.

To assess the trend of correct answers (whether the 'easy' or the 'hard' condition saw an upturning average of correct answers as gameplay continued), a rolling mean was used. This calculated a constant rolling average of correct answers within the finite window of each pair's full set of responses. Subsequently, these means were plotted against the number of turns taken by participants during gameplay, thus an increase in the mean correct responses in each condition demonstrates an upward trend of correct answers as gameplay progressed. For further analysis, I ran a linear mixed model analysis using R Studio where the rolling mean of correct responses was the dependent variable and the fixed effects were condition, turn number, and the interaction between condition and turn number. Each pair of participants was treated as a random effect. This was specified using the following model: RollCorr ~ Condition*TurnNumber + (1|GameID).

To examine whether either Organelle colour or shape was more salient, I compared the colour and shape of the Organelles selected by participants with the correct Organelle for that turn. Separate columns in the dataset were created which binarily indicated if colour and shape matched. Matches were indicated with a '1' in their respective columns, while non-matching colours or shapes were indicated with a '0'. Once these columns were formed, I measured the rolling average of both columns, within the window of each pairs' complete responses. The rolling averages for colour and shape were then plotted individually against the number of turns taken by participants. This indicated which visual element, colour or shape, participants more consistently correctly guessed. From this, a conclusion regarding the saliency of colour and shape could be drawn from the data.

4. <u>Results</u>

Across both the ‘easy’ and ‘hard’ conditions, accuracy was relatively low. In fact, percentage accuracy showed that neither condition was significantly more accurate than the other. The overall percentage accuracy in the ‘easy’ condition was 23.2%, where the lowest individual pair accuracy was 15.4%, and the highest was 28.0%, with a median of 24.0%. In the ‘hard’ condition, the overall percentage accuracy was 24.5%, with the lowest 17.8% and the highest 47.1%, with a median of 20.7%.

However, there was significantly more variation in the number of turns each pair took in the ‘hard’ condition, compared to the ‘easy’ condition. In the ‘easy’ condition, number of turns was relatively consistent, where 65 was the lowest number of turns and 88 was the highest (median of 68). However, in the ‘hard’ condition, number of turns was erratic, with the lowest number of turns taken 51, and the highest 94 (median of 65). As a result, the cumulative frequency of pairs’ points as they accrued throughout gameplay revealed that, on average, those pairs playing in the ‘easy’ condition were scoring points more consistently as number of turns taken increased. The graph in Figure 5 demonstrates the differing consistency with which points were scored in each condition.

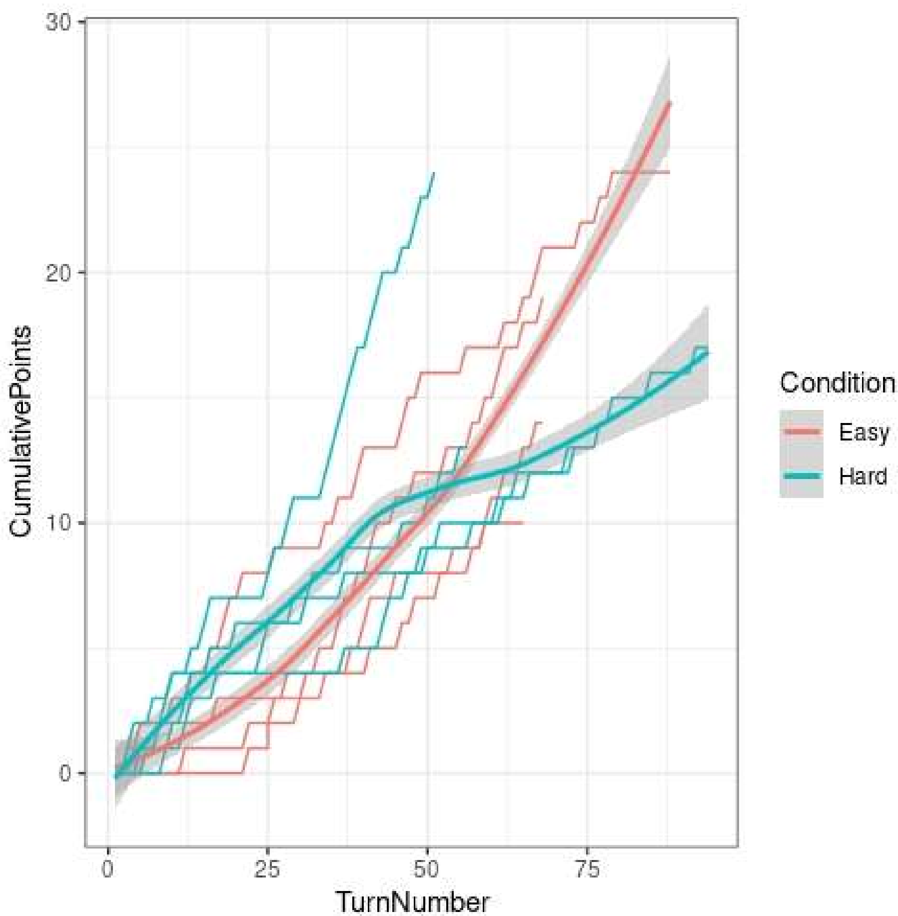


Figure 5: Cumulative points scored across total number of turns for each pair in the 'easy' versus 'hard' conditions

Figure 5 shows the cumulative frequency accrued across both experimental conditions. Between turns 0 and 50, the 'easy' pairs' trend line forms an upwards concave, demonstrating that incidences of point scoring occurred over a greater number of turns than in the 'hard' condition. On average, by turn number 50, the 'easy' pairs had accumulated 10 points. In contrast, the 'hard' pairs' trend line shows more consistent point scoring in the earlier turns of the experiment, earning an average of 10 points by approximately turn 40. However, from turn 50, average points scored by 'easy' participants sharply increases, demonstrating a new consistency of point scoring, which continued until the conclusion of the experiment. In contrast, 'hard' pairs' points scoring slows down considerably, only marginally increasing between turns 40 and 62, before rising steadily until the end of gameplay. Despite a greater continuous consistency of point-scoring demonstrated by the pairs in the ‘easy’ condition, both the ‘easy’ pairs and the ‘hard’ pairs finished gameplay with an average of 17 points.

The one anomalous pair in this study, Hard Pair 3, was excluded from analysis. Despite the pair reaching an almost ceiling level of point scoring over the length of gameplay, the method they employed was, in its totality, derived from their pre-existing knowledge of English. Instead of building consistent mappings between Ferros and Organelles, Hard Pair 3 used Ferro cards to spell out the colour of Organelles using English and laid them down in distinct ways to indicate shape. For example, the Organelle ‘green spiky’ would be communicated by placing five random Ferros in a straight line on the signal sheet, as shown in Figure 6.

Figure 6: Visual image of Hard Pair 3’s English-dependent method of gameplay.

One participant in Hard Pair 3 commented at the end of gameplay that whether Ferros were printed on the cards or not was irrelevant to them, further demonstrating the pair’s disuse on Ferro as a communication system, and their heavy reliance on English. As such, Hard Pair 3’s data was not reflective of emergent structure in a novel language and considerably skewed the remaining data. To mitigate this, the data was omitted from the study and another pair was recruited to complete the experiment in the ‘hard’ condition.

Figure 7 shows the duration, in seconds, of turns taken by each pair in both experimental conditions. A decline in average turn length is observed in both conditions, as expected. However, whilst Figure 7 shows a very similar decline in both conditions, average turn length in the 'easy' condition appears to remain lower than in the 'hard' condition, and decreases gradually and consistently until the end of gameplay. Average length of the 'easy' pairs first turn was 40.35 seconds, decreasing to an average of 18.35 seconds by each pair's final turn. The quickest turn of the experiment occurred in the 'easy' condition, lasting 7.93 seconds and resulting in a point. Despite beginning with an average quicker first turn of 29.81 seconds, the 'hard' pairs' turn length remained plateaued with a marginal increase in turn length from the beginning of gameplay until around turn 40. From turn 40, average turn length in the 'hard' condition declined consistently until the conclusion of gameplay, but remained slower than in the 'easy' condition. The average length of the final turn in the 'hard' condition was 38.27 seconds, 8.46 seconds slower than in the 'easy condition'. The longest turn across both condition was 73.78 seconds and occurred in the 'hard condition'.

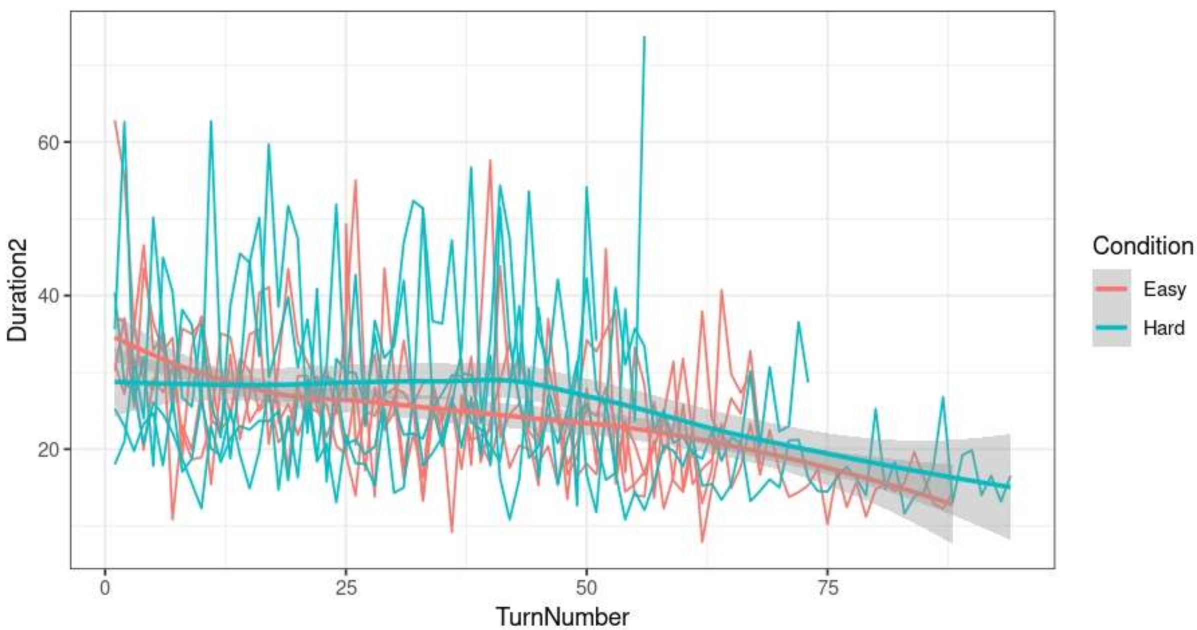


Figure 7: Average duration (s) of each turn taken by participants in both experimental conditions.

In the backdrop of increasingly quicker turn speeds across both conditions and more consistent point scoring in the 'easy' condition, the rolling mean of correct answers was calculated. This involves continuously updating the average of the data (in this case, every 5 turns), including all the data in the set, the output of which is typically a more realistic result than a standard average. Figure 8 shows the rolling mean of both conditions plotted against the number of turns taken by participants. Between turns 25 and 50, it appears that participants in both conditions see an increasing rolling average, indicating that participants are beginning to coordinate with one another. The rolling average for participants in the 'easy' condition raises from 0.17 at turn 25, to 0.29 at turn 50, reflecting an average increase of 0.12 in 25 turns. Meanwhile, in the hard condition, there is a similarly upturning average.

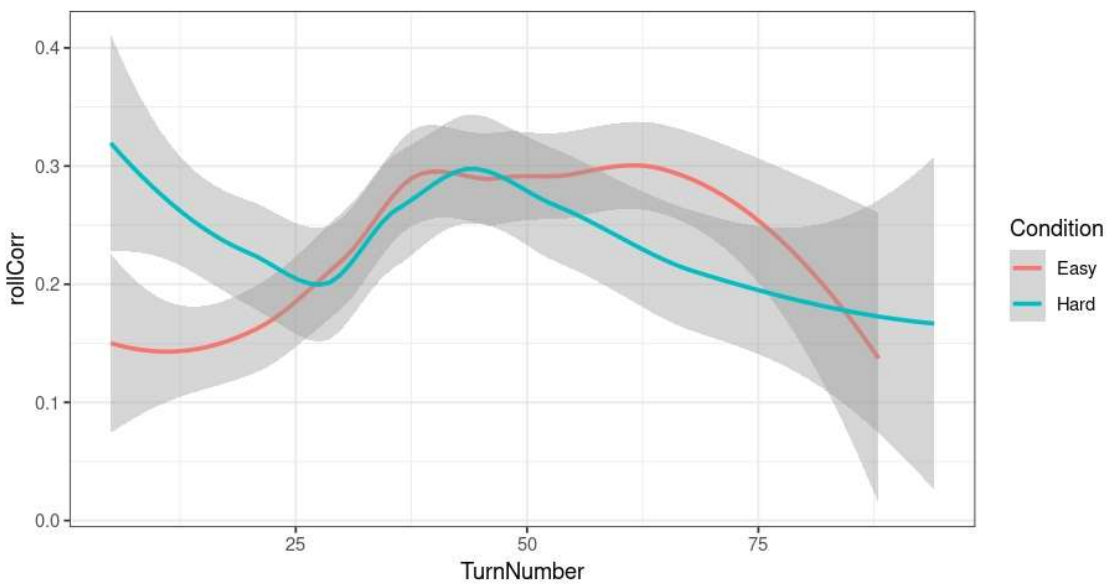


Figure 8: The average rolling mean of points scored across both conditions throughout gameplay

At turn 25, the rolling average of correct answers was 0.2, subsequently raising to 0.3 by turn 40. As turn number increases, Figure 8 shows that the 'easy' condition is beginning to pull away from the hard condition. By turn 60, the average rolling mean in the 'easy' condition remains relatively plateaued, hovering around 0.3. Meanwhile, the rolling average of correct responses in the 'hard' condition begins to consistently decline from turn 40, with a rolling average of 0.24 by turn 60. The data becomes more noisy towards the latter end of Figure 8. This increase in noise coincides with participants beginning to take their final turns leading up to the conclusion of gameplay. There is a noticeable drop off in both conditions, more sharply observed in the 'easy' condition than the 'hard'. This decrease is likely characterised by the sudden drop in the number of groups in each condition still playing. In the later stages of turn taking, some groups begin concluding gameplay, having taken fewer total turns across their 30 minutes of gameplay. In the ‘easy’ condition, there are only 2 pairs left of 4 still playing by turn 75. As such, the dip in Figure 8 reflects this change, particularly in the ‘easy’ condition.

Although Figure 8 indicates that the rolling average of correct answers in the ‘easy’ condition could be pulling away from that of the ‘hard’ condition, the increase of noise in the data casts uncertainty over this conclusion. Statistical analyses present an effective method of quantifying such uncertainty. For this, a linear mixed-effect regression model is used to extract more concrete conclusions from the data, as seen in Table 1.

Random effects			
Groups Name	**Variance**	**Std.Dev.**	
GameID	0.01	0.101	
Residual	0.031	0.176	
Number of obs: 531, groups: GameID, 8			
Fixed effects			
	Estimate	**Std. Error**	**t-value**
(Intercept)	0.1521	0.05567	2.733
ConditionHard	0.0999	0.07863	1.1271
TurnNumber	0.0021	0.00053	3.938
ConditionHard:TurnNumber	-0.0017	0.00075	-2.242
Correlation of Fixed Effects			
	(Intr)		
ConditionHard	-0.708		
TurnNumber	-0.366		
ConditionHard:TurnNumber	0.257		

Table 1: Results from a Linear Mixed-Effects regression using the average rolling mean of correct responses.

Table 1 shows the outcome of the linear mixed-effect regression run on the rolling mean data of points scored. As expected, Turn Number was significant, with a positive estimate value and a t-value of 3.938, indicating that the rolling mean of correct responses was increasing over time in both conditions. However, Table 1 shows that the interaction between Condition and Turn Number was a significant one. Both the negative estimate result and the t-value of -2.625 indicate that the correlation rolling mean is going down. In sum, pairs playing in the hard condition were getting worse over time. This signifies that the correlation rolling mean is not declining in the ‘easy’ condition in the same manner as it is in the 'hard' condition. Over time, those pairs in the 'easy' condition do better than those in the 'hard' condition.

Regardless of condition, there was no observed emergence of compositional structure in pairs' language systems. Gameplay revealed that 7 of 8 participants adopted a holistic approach, communicating using solely 1-1 mappings. As will be further elaborated upon in the Discussion, whilst the remaining pair did adopt a method that involved the production of multiple Ferros, it is doubtful that this reflects a compositional structure. Further discussion regarding the lack of emergent compositional structure is addressed in length in the Discussion.

Another line of investigation concerns the comparison of colour and shape saliency of Organelles. Figures 9 and 10 overwhelming demonstrate that pairs were more accurate guessing shape than colour.

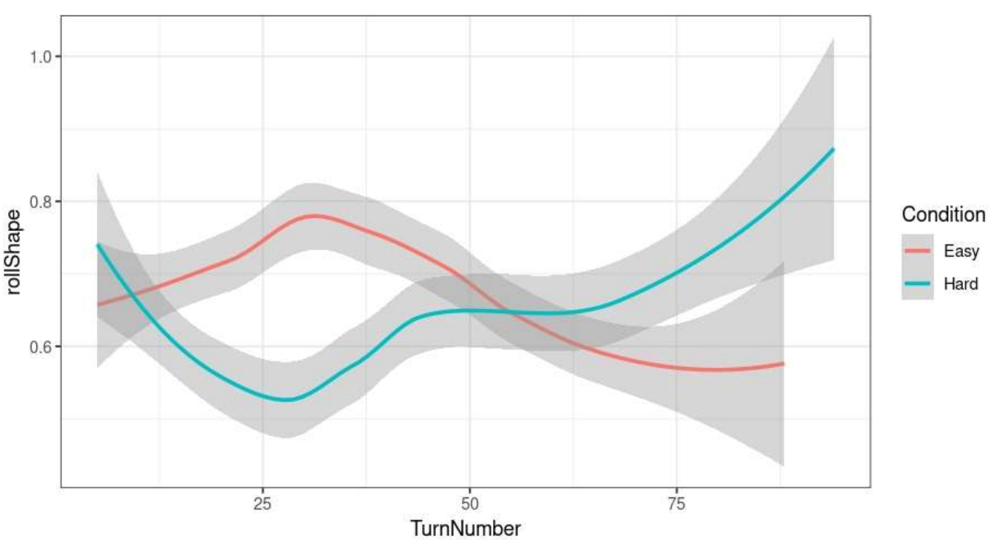


Figure 9: The Rolling Mean of correctly guessed Organelle shape as Turn Number increases in both conditions.

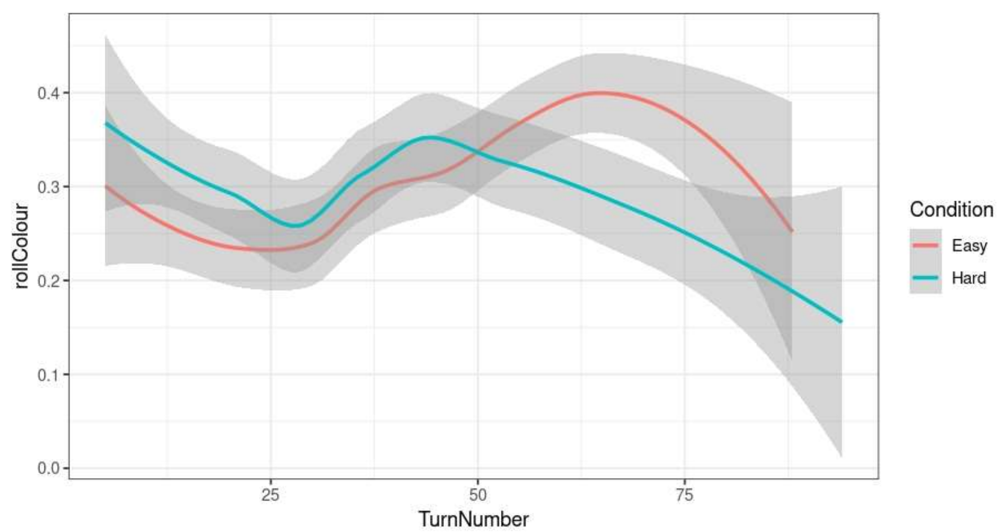


Figure 10: The Rolling Mean of correctly guessed Organelle colour as Turn Number increases in both conditions.

Figure 9 demonstrates that rolling accuracy begins relatively high for shape in both conditions, showing an accuracy of 0.65 in the 'easy' condition and 0.75 in the 'hard' condition. Although immediately declining to an average low of 0.52 by turn 27 in the 'hard' condition, the rolling average picks up steadily, raising to a high of 0.88 by the end of gameplay. Conversely, in the 'easy' condition, the rolling average immediately increases to a high of 0.79 by turn 35. Subsequently, the rolling average begins declining and continues to do so until turn 75, where it sits at 0.58. Before the end of gameplay, there appears to be a subtle upturn which may be indicative of a new increase, but this is cut short by the conclusion of gameplay. It appears that while pairs in the 'easy' condition were more accurate in guessing shape, and as such may have found shape more prominent than colour in the early stages of gameplay, pairs in the 'hard' condition were significantly more accurate in guessing shape towards the latter stages of gameplay.

Figure 10 shows that the rolling average accuracy for colour follows a relatively similar trend in both conditions. Overall, the rolling average for colour remains lower than that of shape in both condition for the duration of gameplay. The rolling average in the 'easy' condition starts low, at 0.3. Immediately, the average declines to a low of 0.23 by turn 25 before picking up and increasing steadily to a high of 0.4 by turn 63. After this, the rolling average drops off in the 'easy' condition until the end of gameplay. In the 'hard' condition, the rolling average starts off similarly at 0.37. Like the 'easy' condition, the rolling average in the 'hard' condition immediately declines to 0.26 by turn 27 before increasing to 0.35 by turn 40. Despite this increase, the rolling average for colour in the 'hard' condition begins declining steadily, dipping as low as 0.20 by the end of gameplay. Whilst shape remained more salient than colour in both conditions despite the fluctuation in accuracy across both variables in both conditions, linear mixed effect regressions provide further insight into the observed trends.

Random effects			
Groups Name	**Variance**	**Std.Dev.**	
GameID	0.004	0.063	
Residual	0.045	0.211	
Number of obs: 531, groups: GameID, 8			
Fixed effects			
	Estimate	**Std. Error**	**t-value**
(Intercept)	0.7476	0.04177	17.899
ConditionHard	-0.1748	0.05886	-2.97
TurnNumber	-0.0016	0.00063	-2.517
ConditionHard:TurnNumber	0.0029	0.00089	3.205
Correlation of Fixed Effects			
	(Intr)		
ConditionHard	-0.71		
TurnNumber	-0.582		
ConditionHard:TurnNumber	0.41		

Table 2: The results of the Linear Mixed-Effect regression run on the average rolling mean of correctly guessed Organelle ***shapes****.*

Table 2 shows the outcome of the linear mixed effects regressions calculated for the rolling mean of guessed Organelle shape. The results show that Condition is significant here with the t-value of -2.970 (Estimate = -0.2748). Similarly, Turn Number was significant, also showing a negative t-value of -2.517 (Estimate = -0.0016). Conversely, the interaction between Condition and Turn Number was of a positive significance here (t-value = 3.205, Estimate 0.0029) indicating that, over time, pairs in the 'hard' condition got better at guessing the shape of Organelles. This result also shows that those in the 'easy' condition do not improve as much over time, which is evident from Figure 8.

Random effects			
Groups Name	**Variance**	**Std.Dev.**	
GameID	0.008	0.088	
Residual	0.037	0.193	
Number of obs: 531, groups: GameID, 8			
Fixed effects			
	Estimate	**Std. Error**	**t-value**
(Intercept)	0.2276	0.05071	4.488
ConditionHard	0.09874	0.07159	1.379
TurnNumber	0.00197	0.00057	3.435
ConditionHard:TurnNumber	-0.00215	0.00082	-2.625
Correlation of Fixed Effects			
	(Intr)		
ConditionHard	-0.708		
TurnNumber	-0.438		
ConditionHard:TurnNumber	0.308		

Table 3: The results of the Linear Mixed-Effect regression run on the average rolling mean of correctly guessed Organelle ***colour****.*

Table 3 shows the results of the linear mixed effect model run on the rolling average of guessed Organelle colour. Unlike in Table 2, Condition is not a significant factor here. However, Turn Number is, with a positive t-value of 3.425 (Estimate -0.00197) making Turn Number more significant for colour than shape in this study. The interaction between Condition and Turn Number is also significant. However, unlike for shape, the negative estimate and t-value (-0.00215 and -2.625 respectively) demonstrates that those pairs playing in the hard condition were getting worse at guessing colour over time. Meanwhile, those playing in the 'easy' condition did not exhibit the same decline in rolling average over time.

5. Discussion

This study asked the question: When presented with a novel language system, how effectively will participants create a communication system between them? In order to investigate this question, I designed a new card game where participants were communicating with a novel language system, Ferro, about a defined meaning space. The emergence of structure in the languages formed by participants were subsequently tracked and observed. I formulated three hypotheses with the aim of answering the research question. These were as follows:

1. Accuracy of participants' turns will increase over the duration of gameplay, in both experiment conditions, as a result of consistent form to meaning mappings between Ferros and Organelles.

2. Compositional structure will emerge in the languages of participants playing in the 'hard' condition given the increased difficulty distinguishing between Ferros leading to constraints on memory.

3. The colour of Organelles will be more salient than the shape and, where compositional structure emerges, reference to colour and shape will follow the perceptual saliency hierarchy (PSH).

A crucial effect of gamification in this study is the lowering of demand characteristics throughout the experiment, resulting in the reduction of unwanted influences on participants' responses (McCambridge et al 2012). Gamification was successful in this regard. Participants were disinterested in behaving like a 'good subject' (McCambridge et al 2012) and were instead immersed in an environment with their partner where they were engaged with the rules of the card game. As a result, participants had a reduced feeling of being watched or observed. Whilst this can result in instances of anomaly, such as Hard Pair 3, we can otherwise be confident that the results of this study, as they are discussed below, reflect the accurate and natural behaviour of participants independent of what I, as the experimenter, wanted to see.

The results of this study demonstrate some important findings. The most significant of these is the relationship between the two experimental conditions and accuracy. Accuracy increased throughout the course of gameplay in the 'easy' condition but not in the 'hard' condition. In fact, results showed that pairs in the 'hard' condition were becoming less accurate over time. Due to the differing distinguishability of Ferro's in both conditions, it is unsurprising that the results indicated that the rolling mean of correct responses in the 'easy' condition was pulling away from the 'hard' condition. Whilst this upward trend in point scoring in the 'easy' condition declined towards the end of gameplay as pairs began to reach then end of their 30 minutes of gameplay, I believe this trend would have continued until pairs reached ceiling. More consistent point scoring was expected in the 'easy' condition due to the ease with which pairs were be able to infer the structure of the meaning space, compared to the 'hard' condition. While accuracy in the 'hard' condition was expected to initially be lower than in the 'easy' condition, a worse average overall performance from 'hard' pairs was not anticipated. Thus, this study's first hypothesis is only somewhat supported.

However, there is an element of uncertainty regarding whether increased accuracy in pairs, and whole conditions, was a direct result of more consistent form-meaning mappings, or just improving coordination between pairs. In a closer look at Hard Pair 5's results, it appeared to show accuracy increasing alongside more consistent form meaning mappings over time. For example, the pair converged on Ferros to express 'green spiky', 'blue blob', 'blue spiky' and 'purple spiky' consistently correctly after only one or two turns communicating each meaning. Therefore, it seems at least in this pair, participants are converging on consistent form to meaning mappings which is ultimately aiding them in solving the task. Though this appears to be the case in Hard Pair 5, a systematic analysis across all pairs in the study would need to be conducted in order to report conclusively whether increasing accuracies were a result of consistent form to meaning mappings or just heightened coordination. Whilst it would have been possible for me to conduct such an analysis, I found that this experiment outputted an incredibly rich dataset with a large amount of variables to be analyzed. To have thoroughly analyzed all of these would not have been possible in this timeframe. As such, I suggest that if this study were to be replicated, an interesting avenue of further research would be to conduct a systematic analysis of pairs' turns to explore whether or not consistent form to meaning mappings have more effect on accuracy than coordination.

Another surprising yet key finding of this study was the lack of emergent compositional structure in pairs in the 'hard' condition. All pairs, aside from Easy Pair 4, adopted a holistic approach, often using 1-1 mappings between forms and meanings. Such holistic mappings were expected in the 'easy' condition as Ferro's distinguishability made inferences about the structure of the form space considerably easier than in the 'hard' condition; the demand on participants' working memories was lower as a result. The only outlier to the holistic approach was Easy Pair 4, who used multiple Ferros to express Organelles in numerous turns throughout gameplay. However, it appears unlikely that this pair was using a compositional structure based on the lack of consistent arrangement and organization of Ferros to communicate Organelles. During gameplay, there were instances where Easy Pair 4 used three Ferros to express an Organelle. In other cases they used two Ferros to express the relevant Organelle. There were even instances where just one Ferro was used, indicating a degree of holistic mapping consistent with the rest of the pairs in the experiment. As a result, it appears that Easy Pair 4 are not an accurate example of the emergence of a compositional structure. The motivation behind the hypothesised emergence of compositional structure in the 'hard' condition was the predicted memory constraint placed upon participants as a consequence of the 'hard' condition's perceptually similar Ferro's. Nevertheless, the results falsified this study's second hypothesis.

The size of the form space relative to meaning space in this experiment could be partly responsible for the lack of emergent compositionality. In this study, there were 10 forms (Ferros) and 10 meanings (Organelles), which undoubtedly encouraged participants to form and converge on simple one-to-one mappings once they were aware that the form and meaning spaces were the same size. This realisation was likely to be advanced further by the experimental set-up, which entailed the entire meaning space being on display in front of participants at all times. This design differs from similar previous experiments (Kirby 2015, Raviv 2019) where the meaning space was not on display in its entirety. A method of potentially mitigating the emergence of simple one-to-one mappings could involve expanding the size of the meaning space gradually, such as in Raviv et al (2019), by adding extra Organelles as gameplay continues. Conversely, the form space could be shrunk so that participants play through the experiment with only 5 Ferros to communicate 10 Organelles. Another method that could be implemented to trigger the emergence of compositionality involves a chain of transmission through participants. This could follow a similar method to the one used in Kirby et al's (2015) study, where participants interacted continuously for a set amount of time before one participant was replaced with someone new. A consequence of manipulating the experimental set-up in these ways is the increase of compressibility pressure on participants during gameplay. Each method adds pressure on participants to make their language more learnable and more simple. In the present study, it was believed that the memory constraint, created as a consequence of the similar Ferro's in the hard condition, would generate sufficient compressibility pressure to trigger compositionality. This was not the case, likely as a result of compressibility pressure not being great enough to encourage participants to converge on a more learnable structure. This is especially true when one-to-one mappings between form and meaning are available, such as in this experiment. This is certainly not to say that a compositional structure cannot emerge in an experiment where participants communicate using Ferro, but that if this study were to be run again, it would be appropriate to manipulate the experimental set up in order to increase compressibility pressure.

It is also worth considering the emergence of compositional structure against the backdrop of previous literature. In much of the previous work where compositionality has emerged, such as in Kirby et al (2015), participants were dealing with forms that followed a CVCV (consonant-vowel-consonant-vowel) structure, which is akin to syllable structure in English. As a result, there is no question that participants will, as long as they are English speakers, implicitly understand the structure of the form space and utilise this immediate understanding effectively within the scope of the experiment. The forms in the present study differ enormously from this structure. When participants look at Ferros, they are looking at a completely novel form space, the structure of which cannot be inferred from their knowledge of a previous language system. While pairs of participants in previous studies would not have experienced difficulty converging on the same interpretation of the structure of the form space, the same cannot be said for the present study. In this experiment, participants had to individually infer the structure of the form space and subsequently converge on the same interpretation with their partner. I believe this underscores the importance of exploring the emergence of linguistic structure using a communication system that is removed from participants' previous knowledge of a language. When presenting participants with forms that consist of CVCV strings, there is the potential that participants are drawing too much on their knowledge of English syllable structure. As a result, the method employed in the current study has a lot of potential relative to earlier work that explores the emergence of linguistic structure in novel communication systems.

A further key finding can be attributed to the comparison between colour and shape saliency. The results showed that shape was considerably more salient than colour throughout the experiment and did not differ depending on which condition participants were playing in. Therefore, the initial part of the third hypothesis has been falsified. This was an unexpected result not least due to the wealth of previous literature that has found colour to be a more powerful cue in visual perception than shape or size (Williams 1966; Turatto and Galfano 2001 cited in Tarenskeen et al 2015). Tarenskeen et al have argued that much of the experimental stimuli in previous literature has presented colour in a "bright and/or highly contrastive" (2015: 3) manner, whereas shape and size were more modestly presented. The perceptual saliency of colour usually makes it stand out against the abstract, geometric figures (triangles, squares etc) usually representing shape in these experimental stimuli. If you look upon a naturalistic scene, colour is likely to be the more visually perceptive cue as abstract shapes, like triangles, are not as frequently occurring in such scenes. However, the context and relationship between colour and shape in the present study varies from previous environments. The meaning space incorporates shapes with more organic forms, such as 'spiky' or 'blob'. So, in the present study, perception of the simple colours used was likely overwhelmed by participants perceiving the more complex naturalistic shapes of the Organelles. Thus, shape was considerably more salient than colour throughout.

The second segment of the this study's third hypothesis predicted that compositionality would display a structure consistent with the PSH. Due to the lack of compositionality emerging in this study, no concrete conclusion can be drawn regarding the structure of compositional signals. Whilst it cannot be overtly stated whether or not participants would produce compositional signals that were congruent with the PSH, I do predict that this would not be the outcome, should this element be explored further. The saliency results in this study show participants' unwavering preference for shape over colour. If a compositional approach were to emerge, it seems logical that shape would retain the same strength of visual perception. As a result, I would expect the PSH to be flouted. Future research could explore further the saliency of colour vs shape in this experiment, ensuring sufficient compressibility pressures are applied to participants in order to trigger the compositional structure within which colour and shape saliency can be analysed further.

6. Conclusion

The present study engaged 9 pairs in a card-based communication game, where participants communicated using a completely novel form space (Ferro) that was totally removed from participants' pre-existing knowledge of a language system. Pairs were split into two groups, 'easy' and 'hard', and the language systems created by each pair were explored for the emergence of linguistic structure. It was expected that pairs in both conditions would become more accurate over time. Furthermore, the challenging task of distinguishing between Ferros in the 'hard' condition underpinned the hypothesis that a compositional structure would emerge within 'hard' pairs' communication. Finally, the colour of meaning cards (Organelles) was expected to be more salient than shape and, where compositional structure emerged, it was hypothesized that this structure would be congruent with the PSH. Results showed that accuracy increased over time in the 'easy' condition especially. The same could not be said for the 'hard' condition, where accuracy initially increased before participants began to get worse over time. Surprisingly, compositional structure did not emerge in any pair in the 'hard' condition, demonstrating that the memory constraint present in this condition did not provide sufficient compressibility pressures to trigger compositionality. Instead, a holistic approach was adopted by all pairs in the experiment, bar one. Finally, results indicated that shape was significantly more salient than colour, despite previous studies showing colour to be more visually perceptive than shape. However, the organic shapes of Organelles differ from the geometric shapes used in previous literature and may be the underlying motivator for the pattern of results in the present study. Were this research to be conducted again, a multitude of alterations to the method have been suggested with the intention that they might produce results that reflect an emergent compositional structure and, as a result, more complex language systems.

List of References

Barrett, J., Skyrms, B. and Cochran, C. (2020). ‘On the Evolution of Compositional Language’. *Philosophy of Science*, 87(5). doi:https://doi.org/10.1086/710367.

Bartlett, F.C. (1932). ‘Remembering: A study in experimental and social psychology’. *British Journal of Educational Psychology*, 3(2), pp.187–192. doi:https://doi.org/10.1111/j.2044-8279.1933.tb02913.x.

Brentari, D. and Goldin-Meadow, S. (2017). ‘Language Emergence’. *Annual Review of Linguistics*, 3(1), pp.363–388. doi:https://doi.org/10.1146/annurev-linguistics-011415-040743.

Caponetto, I., Earp, J. and Ott, M. (2014). ‘Gamification and Education: A Literature Review’. In: C. Busch, ed., *Proceedings of the 8th European Conference on Games Based Learning*. Reading, UK: Academic Conferences and Publishing International Limited.

Christians, G. (2018). ‘The Origins and Future of Gamification’. [online] Available at: https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=1255&context=senior_theses.

Christiansen, M.H. and Chater, N. (2015). ‘The Now-or-Never bottleneck: A fundamental constraint on language’. *Behavioral and Brain Sciences*, 39. doi:https://doi.org/10.1017/s0140525x1500031x.

Cuskley, C. (2019). ‘Alien forms for alien language: investigating novel form spaces in cultural evolution’. *Palgrave Communications*, 5(1). doi:https://doi.org/10.1057/s41599-019-0299-5.

Deterding, S., Dixon, D., Khaled, R. and Nacke, L. (2011). ‘From game design elements to gamefulness: defining ‘gamification’’. In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. [online] pp.9–15. Available at: https://dl.acm.org/doi/abs/10.1145/2181037.2181040.

Downes-Le Guin, T., Baker, R., Mechling, J. and Ruyle, E. (2012). ‘Myths and Realities of Respondent Engagement in Online Surveys’. *International Journal of Market Research*, 54(5), pp.613–633. doi:https://doi.org/10.2501/ijmr-54-5-613-633.

Galke, L. and Raviv, L. (2024). ‘Emergent communication and learning pressures in language models: a language evolution perspective’. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.2403.14427.

Gong, T., Lam, Y.W. and Shuai, L. (2016). ‘Influence of Perceptual Saliency Hierarchy on Learning of Language Structures: An Artificial Language Learning Experiment’. *Frontiers in Psychology*, 7. doi:https://doi.org/10.3389/fpsyg.2016.01952.

Hamari, J., Koivisto, J. and Sarsa, H. (2014). ‘Does Gamification Work? -- a Literature Review of Empirical Studies on Gamification’. *2014 47th Hawaii International Conference on System Sciences*, [online] pp.3025–3034. doi:https://doi.org/10.1109/hicss.2014.377.

Huber, S.E., Cortez, R., Kiili, K., Lindstedt, A. and Ninaus, M. (2023). ‘Game elements enhance engagement and mitigate attrition in online learning tasks’. *Computers in Human Behavior*, [online] 149. doi:https://doi.org/10.1016/j.chb.2023.107948.

Huotari, K. and Hamari, J. (2016). ‘A definition for gamification: anchoring gamification in the service marketing literature’. *Electronic Markets*, [online] 27(1), pp.21–31. doi:https://doi.org/10.1007/s12525-015-0212-z.

Kirby, S., Cornish, H. and 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), pp.10681–10686. doi:https://doi.org/10.1073/pnas.0707835105.

Kirby, S., Griffiths, T. and Smith, K. (2014). ‘Iterated learning and the evolution of language’. *Current Opinion in Neurobiology*, 28(1), pp.108–114. doi:https://doi.org/10.1016/j.conb.2014.07.014.

Kirby, S., Tamariz, M., Cornish, H. and Smith, K. (2015). 'Compression and communication in the cultural evolution of linguistic structure'. *Cognition*, 141, pp.87–102. doi:https://doi.org/10.1016/j.cognition.2015.03.016.

Koivisto, J. (2017). 'Gamification : A study on users, benefits and literature'. [online] Available at: https://trepo.tuni.fi/bitstream/handle/10024/102270/978-952-03-0550-5.pdf?sequence=7&isAllowed=y [Accessed 2024].

Krath, J., Schürmann, L. and von Korflesch, H.F.O. (2021). 'Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning'. *Computers in Human Behavior*, 125(1), p.106963. doi:https://doi.org/10.1016/j.chb.2021.106963.

Lee, J. and Hammer, J. (2011). *(PDF)* 'Gamification in Education: What, How, Why Bother?' [online] ResearchGate. Available at: https://www.researchgate.net/publication/258697764_Gamification_in_Education_What_How_Why_Bother.

McCambridge, J., de Bruin, M. and Witton, J. (2012). 'The effects of demand characteristics on research participant behaviours in non-laboratory settings: A systematic review'. *PLoS ONE*, 7(6), p.e39116. doi:https://doi.org/10.1371/journal.pone.0039116.

Mcgonigal, J. (2012). 'Reality is broken: why games make us better and how they can change the world'. London: Vintage.

Morschheuser, B., Hamari, J. and Koivisto, J. (2016). 'Gamification in Crowdsourcing: A Review'. [online] IEEE Xplore. doi:https://doi.org/10.1109/HICSS.2016.543.

Muthiyan, G., Kasat, P.A, Vij, V., Solanki, R., C Kirubhanand and Bharat Sontakke (2023). 'Effectiveness of an Innovative Card Game as a Supplement for Teaching Factual Content to Medical Students: A Mixed Method Study'. *Cureus*. doi:https://doi.org/10.7759/cureus.47768.

Orne, M.T. (1996). 'Demand Characteristics'. *Introducing Psychological Research*, pp.395–401. doi:https://doi.org/10.1007/978-1-349-24483-6_59.

Pedale, T., Mastroberardino, S., Capurso, M., Macrì, S. and Santangelo, V. (2022). ‘Developmental differences in the impact of perceptual salience on short-term memory performance and meta-memory skills’. *Scientific Reports*, [online] 12(1), p.8185. doi:https://doi.org/10.1038/s41598-022-11624-8.

Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C. and Cena, F. (2018). ‘Strengthening gamification studies: Current trends and future opportunities of gamification research’. *International Journal of Human-Computer Studies*, 127. doi:https://doi.org/10.1016/j.ijhcs.2018.11.007.

Ratinho, E. and Martins, C. (2023). ‘The role of gamified learning strategies in student’s motivation in high school and higher education: A systematic review’. *Heliyon*, [online] 9(8), p.e19033. doi:https://doi.org/10.1016/j.heliyon.2023.e19033.

Raviv, L., Meyer, A. and Lev-Ari, S. (2019). ‘Compositional structure can emerge without generational transmission’. *Cognition*, 182, pp.151–164. doi:https://doi.org/10.1016/j.cognition.2018.09.010.

Ryan, R.M. and Deci, E.L. (2000). ‘Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being’. *American Psychologist*, 55(1), pp.68–78.

Sailer, M., Hense, J., Mandl, H. and Klevers, M. (2013). ‘Psychological Perspectives on Motivation through Gamification’. *Interaction Design and Architecture(s)*, (19), pp.28–37. doi:https://doi.org/10.55612/s-5002-019-002.

Sailer, M., Hense, J.U., Mayr, S.K. and Mandl, H. (2017). ‘How Gamification motivates: an Experimental Study of the Effects of Specific Game Design Elements on Psychological Need Satisfaction’. *Computers in Human Behavior*, [online] 69(69), pp.371–380. doi:https://doi.org/10.1016/j.chb.2016.12.033.

Saldana, C., Kirby, S., Truswell, R. and Smith, K. (2019). ‘Compositional Hierarchical Structure Evolves through Cultural Transmission’: An Experimental Study. *Journal of Language Evolution*, 4(2), pp.83–107. doi:https://doi.org/10.1093/jole/lzz002.

Selten, R. and Warglien, M. (2007). 'The emergence of simple languages in an experimental coordination game'. *Proceedings of the National Academy of Sciences*, 104(18), pp.7361–7366. doi:https://doi.org/10.1073/pnas.0702077104.

Smith, K., Kirby, S. and Brighton, H. (2003). 'Iterated Learning: A Framework for the Emergence of Language'. *Artificial Life*, 9(4), pp.371–386. doi:https://doi.org/10.1162/106454603322694825.

Tarenskeen, S., Broersma, M. and Geurts, B. (2015). 'Overspecification of color, pattern, and size: salience, absoluteness, and consistency'. *Frontiers in Psychology*, 6. doi:https://doi.org/10.3389/fpsyg.2015.01703.

Thurairasu, V. (2022). 'Gamification-Based Learning as The Future of Language Learning: An Overview'. *European Journal of Humanities and Social Sciences*, 2(6), pp.62–69. doi:https://doi.org/10.24018/ejsocial.2022.2.6.353.

Turatto, M. and Galfano, G. (2001). 'Attentional capture by color without any relevant attentional set'. *Perception & Psychophysics*, 63(2), pp.286–297. doi:https://doi.org/10.3758/bf03194469.

Werbach, K. and Hunter, D. (2012). 'For the win : how game thinking can revolutionize your business'. Philadelphia: Wharton Digital Press.

Williams, L.G. (1966). 'The effect of target specification on objects fixated during visual search'. *Perception & Psychophysics*, 1(9), pp.315–318. doi:https://doi.org/10.3758/bf03215795.

Appendices

Appendix 1

Participant Information Sheet

Project	Communication-based Card Game
Principal Investigator	Sophie Woodman and Dr Christine Cuskley
Researcher collecting data	Sophie Woodman

What is this document? This document explains what kind of study you are being asked to take part in, what your rights are, and what will be done with your data. If there are any special benefits or risks to participation, they will be explained here. Please read the information carefully and retain it for your records.

The Project: You are about to participate in a study which involves playing a card-based communication game. You will play this game in pairs and will be recorded using a video camera. Your session should last for about 40 minutes. You will be given full instructions before you begin.

Risks and benefits: There are no anticipated risks to participating in this study. The only benefits to you personally are those you draw from making a contribution to our knowledge about how people understand and use language.

Right to Withdraw: Your participation is **voluntary**, and you may **stop at any time for any reason without penalty**. To withdraw during participation, please contact Sophie Woodman via email (s.e.woodman2@newcastle.ac.uk) and express your desire to withdraw from the study. If you withdraw, any data that you have already provided or produced will be deleted. To withdraw after your participation is complete, please contact Sophie Woodman via email (s.e.woodman2@newcastle.ac.uk) and express your desire to withdraw from the study. If you choose to withdraw, all your data will be deleted. You may withdraw your data **up to the date of completion of the project**, which is planned to be on 01/05/2024.

Confidentiality: We will not be collecting any identifying information, and none of your responses can be associated with your name or with any other personal details.

Your Data: Any data you provide will be handled in accordance with the UK Data Protection Act of 2018. Your anonymous data will be securely stored and may be used only by the above-named researcher(s) as well as by other qualified researchers for teaching or research purposes, and in professional presentations. Any identifying information you have provided during recruitment or on consent forms will only be accessible by the named researchers, stored separately from your response data, and not used as part of the research. After the study's completion date (planned to be on 01/05/2024) any identifying information will be deleted. Anonymised video recordings may be played in part or in full by the researcher, but only in the context of academic assessments.

Contact Information: This research is being conducted by Sophie Woodman at Newcastle University and overseen by Dr Christine Cuskley. The researcher can be contacted via email (s.e.woodman2@newcastle.ac.uk), or phone (+44 7963461493) for questions or to report a research-related problem. If you have any concerns regarding your rights as a participant in this research, you can contact the ethics convenor of the School of English Literature, Language and Linguistics of Newcastle University (currently Dr William van der Wurff, w.a.m.van-der-wurff@ncl.ac.uk).

Appendix 2

1	yellow_spiky	26	blue_spiky	51	yellow_blob	76	red_spiky
2	red_spiky	27	purple_blob	52	blue_spiky	77	green_spiky
3	purple_spiky	28	green_blob	53	purple_spiky	78	red_blob
4	blue_blob	29	yellow_blob	54	yellow_spiky	79	purple_blob
5	red_blob	30	red_spiky	55	purple_blob	80	yellow_blob
6	green_blob	31	yellow_spiky	56	green_blob	81	blue_blob
7	yellow_blob	32	yellow_blob	57	red_spiky	82	red_blob
8	blue_spiky	33	blue_blob	58	red_blob	83	yellow_spiky
9	purple_blob	34	red_blob	59	green_spiky	84	purple_blob
10	green_spiky	35	green_spiky	60	blue_blob	85	purple_spiky
11	red_spiky	36	blue_spiky	61	red_spiky	86	green_spiky
12	blue_spiky	37	purple_spiky	62	green_spiky	87	red_spiky
13	yellow_spiky	38	red_spiky	63	blue_spiky	88	blue_spiky
14	red_blob	39	purple_blob	64	blue_blob	89	yellow_blob
15	purple_blob	40	green_blob	65	purple_blob	90	green_blob
16	green_spiky	41	purple_blob	66	purple_spiky	91	red_spiky
17	yellow_blob	42	yellow_spiky	67	red_blob	92	purple_blob
18	blue_blob	43	red_blob	68	green_blob	93	green_spiky
19	purple_spiky	44	purple_spiky	69	yellow_spiky	94	red_blob
20	green_blob	45	green_blob	70	yellow_blob	95	purple_spiky
21	yellow_spiky	46	red_spiky	71	green_blob	96	blue_spiky
22	green_spiky	47	blue_spiky	72	yellow_spiky	97	green_blob
23	blue_blob	48	green_spiky	73	purple_spiky	98	yellow_spiky
24	red_blob	49	yellow_blob	74	blue_blob	99	blue_blob
25	purple_spiky	50	blue_blob	75	blue_spiky	100	yellow_blob