Languages adapt to their contextual niche

James Winters[*], Simon Kirby, Kenny Smith

School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, United Kingdom

[*]Corresponding author: J.R.Winters@sms.ed.ac.uk

Abstract

It is well established that context plays a fundamental role in how we learn and use language. Here we explore how context links short-term language use with the long-term emergence of different types of language systems. Using an iterated learning model of cultural transmission, the current study experimentally investigates the role of the communicative situation in which an utterance is produced (SITUATIONAL CONTEXT) and how it influences the emergence of three types of linguistic systems: UNDERSPECIFIED languages (where only some dimensions of meaning are encoded linguistically), HOLISTIC systems (lacking systematic structure) and SYSTEMATIC languages (consisting of compound signals encoding both category-level and individuating dimensions of meaning). To do this, we set up a discrimination task in a communication game and manipulated whether the feature dimension shape was relevant or not in discriminating between two referents. The experimental languages gradually evolved to encode information relevant to the task of achieving communicative success, given the situational context in which they are learned and used, resulting in the emergence of different linguistic systems. These results suggest language systems adapt to their contextual niche over iterated learning.

1) Introduction

One of the fundamental axioms of modern cognitive-functional linguistics is that "[word] meaning is highly context-sensitive, and thus mutable" (Evans, 2005: 71). When interpreting a particular utterance, language users must not only rely on the meaning encoded in linguistic forms, but also on what they infer from contextual information. Such notions were explicitly acknowledged in the early work of Grice (1957), with a distinction being made between

SIGNAL MEANING[1] and CONTEXTUAL MEANING (Evans & Green, 2006). Signal meaning refers to the senses stored in semantic memory, forming part of the user's linguistic knowledge. Contextual meaning is constructed *on-line* and constitutes an extension of the original signal meaning through an individual's inferential capacities (cf. Evans & Green, 2006; Hoefler, 2009: 6). Put simply: "[...] some meaning is encoded in linguistic forms and some is inferred" (Wedgwood, 2007: 652).

In this sense, context broadly refers to the set of premises used in interpreting an utterance, besides the information already specified in the signal meaning, and constitutes a psychological construct that comprises a subset of an individual's assumptions about the world (Sperber & Wilson, 1986: 15-16)[2]. Consider the word MOLE. Besides referring to a small burrowing animal, MOLE can also denote a form of espionage, a type of birthmark and a unit in chemistry. Each of these senses are said to be stored in semantic memory, with their use and interpretation being governed by the very specific contexts in which they occur. Viewed in isolation words such as MOLE might be construed as communicatively dysfunctional. Yet, in context, it is typically easy to distinguish one sense from another. Having specific knowledge of the context thus enables a hearer to change their expectations regarding the intended meaning of a given word. In other words, when the context is known and informative, it necessarily decreases uncertainty (Piantadosi, Tily & Gibson, 2012).

As context is used as a resource to reduce uncertainty, it might alter our conception of how an optimal communication system should be structured (Zipf, 1949; Piantadosi, Tily & Gibson, 2012). Levinson (2000: 29), for instance, argues that our cognitive abilities favour

---

[1] We use *signal meaning* to refer to what Evans & Green (2006) refer to as *encyclopaedic meaning*.
[2] This can refer to the wider *sentential context* in which a word is embedded as well as the *situational* and *interpersonal* contexts that make up the salient common ground, among others (see: Bach, 2012; Evans & Green, 2006: 221).

communication systems which are skewed in their design towards hearer inference over speaker effort. Meanwhile, Pinker & Bloom (1990) note language exhibits design for communication because it allows for "minimising ambiguity *in context*" (pg. 713, emphasis added). Evidence for the role of context is also apparent in the way we structure our utterances, with syntax being sensitive to the wider discourse and the immediate communicative needs of interlocutors (Chafe, 1976; DuBois, 1987; Fery & Krifka, 2008). Furthermore, these immediate communicative needs can give rise to longer-term patterns: here, the way in which speakers pragmatically design utterances (INVITED INFERENCES, Traugott & Konig, 1991), as well as how hearers interpret utterances (CONTEXT-INDUCED INTERPRETATION, Heine, Claudi & Hunnemeyer, 1991), is posited to play a fundamental role in historical processes, such as grammaticalization (cf. Traugott & Trousdale, 2013).

There are a number of different kinds of context we could talk about in relation to a particular usage event (Evans & Green, 2006; Bach, 2012). Our present study is specifically focused on the SITUATIONAL CONTEXT: the immediate communicative environment in which an utterance is situated (Evans & Green, 2006: 221) and how it influences the distinctions a speaker needs to convey. In an experimental setting, situational context can be manipulated by tailoring both the types of stimuli and the way in which they are organised. For example, in a study examining how adjectives were used in referring expressions, Sedivy (2005) discovered that speakers were more likely to use an adjective when one object shared a feature dimension with another object (e.g., a blue cup and green cup), but not when the object belonged to a different category (e.g., a cup and a teddy bear). Similarly, Ferreira, Slevc & Rodgers (2005) found that when speakers were faced with conceptual ambiguities, such as having to discriminate between two types of bat (the flying mammal), they would disambiguate on a relevant dimension (e.g., using *the small bat* in their utterance rather than

just *the bat* when a large bat was also present in the context), whereas when speakers were presented with linguistic ambiguities (e.g., a baseball bat and an animal bat) they were less likely to engage in ambiguity avoidance.

If the situational context plays a fundamental role in how language is structured, then the general observation that *some meaning is encoded* and *some is inferred* leaves open the questions: (i) To what extent does the situational context influence the encoding of features in the linguistic system? (ii) How does the effect of the situational context work its way into the structure of language? To help answer these questions we investigate how situational context influences the emergence of linguistic systems. Using an artificial language paradigm, we experimentally simulate cultural transmission in a pair-based communication game setup (cf. Scott-Phillips & Kirby, 2010; Galantucci, Garrod & Roberts, 2012). Participants learn an artificial language which provides labels for a set of pictures, 'meanings' to be communicated. These stimuli vary on the dimension of shape, with each referent also having a unique, idiosyncratic element. After learning the language, participants play a series of communication games with their partner, taking turns to describe pictures for each other. We modified the situational context in which communication took place by manipulating whether the feature dimension of shape was relevant or not for a discrimination task: for example, some participants would encounter only situational contexts in which the objects to be discriminated during communication differed in shape, whereas others would be confronted with contexts in which the objects to be discriminated during communication were of the same shape. Finally, these pairs of participants were arranged into transmission chains (Kirby, Cornish & Smith, 2008; Scott-Phillips & Kirby, 2010; Thiesen-White, Kirby & Oberlander, 2011), such that the language produced during communication by the *n*th pair in a chain became the language that the *n+1*th pair attempted to learn. This method allows us to

investigate how the artificial languages change and evolve as they are adapted to meet the participants' communicative needs and/or as they are passed from individual to individual via learning. We predict that languages in different types of situational context will adapt to become optimally structured as follows:

- When the feature dimension of shape always differs between pairs of referents which are to be discriminated, we predict that the languages will evolve to only encode shape in the linguistic signal, and become underspecified on all other dimensions.

- When the feature dimension of shape is always shared between pairs of referents which are to be discriminated, we predict that a holistic systems will emerge, in which each referent is associated with an idiosyncratic label that encodes that referent's idiosyncratic feature;

- When the feature dimension of shape sometimes differs and is sometimes shared within pairs of referents, we predict that the languages will become systematically structured to encode both the shape (via a category marker) and idiosyncratic features (via an individuating element of the signal).

1.1   Iterated Learning and Communication Games: A method for investigating the emergence and evolution of language

Language is not only a conveyer of cultural information, but is itself a socially learned and culturally transmitted system, with an individual's linguistic knowledge being the result of observing and reconstructing the linguistic behaviour of others (Kirby & Hurford, 2002). This process can be explored experimentally using ITERATED LEARNING: a cycle of continued production and induction where individual learners are exposed to a set of data, which they must then reproduce and pass on to the next generation of learners (Kirby, Cornish & Smith, 2008).

Using this method, researchers have demonstrated that cultural transmission can account for the emergence of some design features in language, including ARBITRARINESS (Thiesen-White, Kirby & Oberlander, 2011; Caldwell & Smith, 2012), REGULARITY (Reali & Griffiths, 2009; Smith & Wonnacott, 2010), DUALITY OF PATTERNING (Verhoef, 2012) and SYSTEMATIC COMPOSITIONAL STRUCTURE (Kirby *et al.*, 2008; Theisen-White *et al.*, 2011). Typically, a participant is trained on a target system (e.g., an artificial language) and then tested on their ability to reproduce what they have learned, with the test output being used as the training input for the next participant in a chain.

These studies show that cultural transmission can account for the emergence of structure in communication systems. In particular, communication systems adapt to constraints inherent in the learning process: domain-general limitations in our memory and processing capabilities (Christiansen & Chater, 2008) introduce a LEARNABILITY PRESSURE (Brighton, Kirby & Smith, 2005), meaning that languages that are difficult to learn tend not to be accurately reproduced, and therefore change. Recent work in this paradigm shows that the incorporation of situational context can change the extent to which the evolving language encodes certain features of referents. Silvey, Kirby & Smith (2014) show, using a transmission chain paradigm, that word meanings evolve to selectively preserve distinctions which are salient during word learning. Using a pseudo-communicative task, where participants needed to discriminate between a target meaning and a distractor meaning, the authors were able to manipulate which meaning dimensions (SHAPE, COLOUR and MOTION) were relevant and irrelevant in conveying the intended meaning. If a meaning dimension was backgrounded, in that it was not relevant in distinguishing between the target and distractor, then the languages evolved not to encode this particular meaning dimension. Instead, the languages converged

on underspecified systems based on the relevant feature dimensions for discriminating between meanings.

However, language is not merely a task of passively remembering and reproducing a set of form-meaning pairings. Language is also a process of JOINT ACTION (Bratman, 1992; Clark, 1996; Croft, 2000): that is, language is fundamentally a social and interactional phenomenon, whereby the role of usage, communication and coordination are salient pressures on the system (also see: Tomasello, 2008; Bybee, 2010). Experimental communication games have been used to investigate the emergence of combinatorial (Galantucci, Kroos & Rhodes, 2010) and compositional (Selten & Warglein, 2007) structure, the emergence of arbitrary symbols from iconic signs (Garrod, Fay, Lee, Oberlander & MacLeod, 2007), and how common ground influences the extent to which a communication can become established in the first place (Scott-Phillips, Kirby & Ritchie, 2009).

Converging evidence from iterated learning and communication games point to both learning and communication as powerful forces in shaping the structure of language (Smith, Tamariz & Kirby, 2013; Fay & Ellison, 2013). With this in mind, the basic premise of the current experiment is to expend upon this work by: (a) adding a communicative element to the experimental setup of Silvey, Kirby & Smith (2014), and (b) manipulating the types of situational context.

1.2   The Problem of Linkage: Language Strategies and the emergence of language systems

Explaining how context works its way into the structure of language requires that we consider the PROBLEM OF LINKAGE (Kirby, 1999; Kirby, 2012). Rather than there being a straightforward link between our individual cognitive machinery and the features we observe

in language, we are instead faced with an additional dynamical system: SOCIO-CULTURAL TRANSMISSION. Treating language as a COMPLEX ADAPTIVE SYSTEM (Beckner et al., 2009; Cornish, Tamariz & Kirby, 2009) solves this problem of linkage because we can consider how short-term LANGUAGE STRATEGIES (Evans & Green, 2006: 110) used in solving immediate communicative needs can give rise to LANGUAGE SYSTEMS through long-term patterns of learning and use (Bleys & Steels, 2009; Steels, 2012; Beuls & Steels, 2013).

The language strategy a speaker selects to enable a listener to identify their intended meaning is dependent not only on the referential information available, but also the context in which the utterance is situated. Take the relatively simple communicative situation in Figure 1: here, there are several language strategies that a language user could employ to convey the intended meaning. In context 1A, the intended meaning can easily be conveyed by using the label DOG as opposed to CAT. If, however, the situational context pairs the intended referent with another dog (as in context 1B), then it makes little sense to use the referential label of DOG, as the listener is very unlikely to be able to distinguish between the two referents on the basis of that label. Instead, other strategies must be employed, such as providing a unique identifier that is more specialised (DALMATIAN) or creating a compound signal (Ay, Flack & Krakauer, 2007) that has both specialised and generalised components (SPOTTED DOG).

[Figure One goes here]

The current experiment explores how these short-term strategies of achieving communicative success in a situational context influence the emergence of different types of language systems. In particular, we focus on the evolution of three types of language systems: UNDERSPECIFIED, HOLISTIC and SYSTEMATIC. Underspecification captures the observation

that languages abstract across referents by encoding some feature dimensions and ignoring others (Silvey, Kirby & Smith, 2014). Using the examples above, the word DOG is underspecified with respect to whether or not its referent is spotted or brown. Conversely, the labels DALMATIAN, POODLE, SIAMESE and TABBY are holistic, in that they embody an arbitrary set of one-to-one mappings between signals and their meanings[3]: holistic signals serve the purpose of *individuation* (Lyons, 1977). Finally, in a systematic mapping between forms and meanings the signals share elements of form (unlike in a holistic mapping, where each signal is unrelated to the other signals) but are nonetheless one-to-one: systematic languages consist of compound signals (e.g., SPOTTED DOG), whereby part of the structure refers to a general-level category (e.g., DOG) and part of the structure refers to an individuating component (e.g., SPOTTED).

To test for the effect of situational context, we use a GUESSING GAME setup (cf. Steels, 2003; Silvey, Kirby & Smith, 2014): the task is to discriminate between pairings of a target object and a distractor object. In our case, possible referents are drawn from a set of images which vary in shape (see Figure 2 below). Manipulating these pairings gives us three experimental conditions based on: (a) whether the feature dimension of shape is relevant or not in discriminating between two referents, and (b) the extent to which stimuli pairings remain consistent over time with respect to the relevance of the feature dimension of shape.

[Figure Two goes here]

---

[3]It should be noted that these holistic labels are also underspecified in respect to other possible signals (e.g., *dalmatian puppy*, *spotted dalmatian*). The important point to keep in mind is the relevance of dimension we are trying to signal. In this case, the signal *dog* is underspecified when compared with *dalmatian* and *poodle*. We could also highlight instances where *dog* is more specific than other signals, such as *animal*. Such notions are widely acknowledged in any linguistic theory that takes into account hierarchical relations between referents.

In the SHAPE-DIFFERENT condition, pairings of target and distractor are constructed such that the feature dimension of shape is always relevant with respect to discrimination, i.e. target and distractor differ in shape. Since the two objects in such situational contexts have different shapes, then they can be discriminated merely by referring to shape. We therefore predict that the languages in the Shape-Different condition will evolve to become underspecified, specifying shape but not differentiating between the objects within a given shape category: such an underspecified system is functionally adequate for achieving communicative success in this situational context, and is highly learnable (Kirby, Cornish & Smith, 2008). Conversely, in the SHAPE-SAME condition target and distractor are always of the same shape – differing only on their idiosyncratic features. Consequently, the feature dimension of shape is always irrelevant in discrimination, and therefore does not need to be specified linguistically, with abstracting across referents of the same shape being communicatively dysfunctional for these situational contexts. We therefore predict that holistic systems will emerge in the Shape-Same condition, where each individual referent is associated with a unique label that maps onto its idiosyncratic feature. Lastly, for the MIXED condition we manipulated the predictability of situational contexts across trials: on some trials target and distractor share the same shape and on others they differ in shape. When encountering this mix of situational contexts, we hypothesise languages will become systematically structured, encoding in the linguistic signal both the basic-level of shape and individuating information of the idiosyncratic feature. Furthermore, we expect that the labels for the basic-level feature will becomes conventionalised earlier than those specifying the individuating information, with participants attempting to meet their immediate communicative needs on a piecemeal basis, through minimising effort and maximising communicative success; the quickest way to achieve this would be for participants to first align on conventional forms for two shapes (as this minimises effort and will ensure communicative success in contexts where shape is

relevant in discrimination) followed by conventional forms for the eight idiosyncratic features (as these are needed to make these distinctions in contexts where shape is irrelevant in discrimination).

1.3   Ecologically-Sensitive, Learning Bias and Historically Contingent Accounts

Our prediction that manipulations to the situational context will bias the probability of one linguistic system emerging over another is consistent with a broader class of predictions that we will term ECOLOGICALLY-SENSITIVE accounts. Under this perspective, languages adapt to the structure of their niche in an analogous manner to that of biological organisms: just as environmental niches constrain and guide the evolution of species, so too are socio-cultural niches salient constraints on the types of languages that emerge (Lupyan & Dale, 2010). The ecologically-sensitive account is consistent with a range of observations including: that social structure patterns with differences in language structure (Wray & Grace, 2007; Lupyan & Dale, 2010); that word frequency is a product of the range of individuals and topics (Altmann, Pierrehumbert & Motter, 2011); that interactional constraints and conversational infrastructure lead to cultural convergence of linguistic form (Dingemanse, Torreira & Enfield, 2013); that objects and events in the world guide word learning discrimination (Ramscar, Yarlett, Dye, Denny & Thorpe, 2010); that word length patterns with the complexity of the meaning space (Lewis, Sugarman & Frank, 2014); that the structure of languages is shaped by the structure of meanings to be communicated (Perfors & Narravo, 2014).

These ecologically-sensitive accounts can be contrasted with two other theoretical perspectives that make different predictions about the relationship between the situational context and the emergence of linguistic systems. The first of these is the LEARNING BIAS

approach. This makes the prediction that language structure is closely coupled to the prior expectations and biases of language learners (e.g. Griffiths & Kalish, 2007; Reali & Griffiths, 2009; Fedzechkina, Jaeger & Newport, 2012; Culbertson, Smolensky & Wilson, 2013; Culbertson & Adger, 2014). The learning bias approach can be further contrasted with what we term the HISTORICAL CONTINGENCY account, which holds that the types of systems that emerge are primarily constrained by random historical events, subtly biasing the language in one direction or another. When compared with the ecologically-sensitive and the learning bias accounts, a historical contingency prediction is that language structure is the result of lineage-specific outcomes (Lass, 1997), with "the current state of a linguistic system shaping and constraining future states" (Dunn *et al*. 2011: 1).

In their extreme incarnations, the learning bias and historical contingency accounts both predict that manipulating the situational context will have little effect on the types of systems that emerge in our experiment. For a learning bias account we would predict considerable convergence across all experimental conditions: there will be a globally-optimal solution in terms of a prior constraint (or set of constraints), with the languages then converging towards this prior. By contrast, the historical contingency account would predict a much higher degree of variation in the types of systems that eventually emerge, with the states of these systems being better predicted by individual variation and lineages than by either contextual or prior cognitive constraints.

2 Method

2.1 Participants

72 undergraduate and graduate students at the University of Edinburgh (42 female, median age 22) were recruited via the SAGE careers database and randomly assigned into 12 diffusion chains. Each chain consisted of a pair of initial participants who learned a random language, and two pairs of successive participants who learned the previous pair of participants' output language, making 3 generations in total. These chains were further subdivided into three experimental conditions (see §2.3).

2.2   Stimuli: Images and Target Language

Participants were asked to learn and then produce an alien language, consisting of lowercase labels paired with images. The images were drawn from a set of 8 possible pictures, which varied on the dimension of shape (4 blobs and 4 stars), with each individual image also having 1 unique, idiosyncratic subordinate element (see Figure 2).

The training language for the first participant pair in each chain was created as follows. From a set of vowels (a,e,i,o,u) and consonants (g,h,k,l,m,n,p,w) we randomly generated 9 CV syllables which we then used to randomly generate a set of 24 2-4 syllable words. These parameters ensured that there were 3 unique labels for every picture. Each chain was initialised with a different random language. The training language for later pairs of participants consisted of the language produced by the previous participant pair while communicating (see below).

2.3   Procedure: Training Phase and Communication Phase

At the start of the experiment, participants were told they would first have to learn and then communicate using an alien language. Participants completed the experiment in separate booths on networked computers. The experiment consisted of two main phases: a TRAINING PHASE and a COMMUNICATION PHASE. Before each phase began, participants were given detailed information on what that phase would involve and were explicitly told not to use English or any other language they knew during the experiment[4]. For the training phase, participants were trained separately, and it was only during the communication phase that they interacted (remotely, over the computer network).

2.3.1   Training Phase

In each training trial, the participant was presented with a label and two images, one of which was the target and one a distractor. The participant was told that the alien wanted them to pick which of the two images corresponded to the label. Once the participant had selected an image (by clicking on it using the mouse) they were told whether their choice was correct or incorrect, shown the label and target image for 2 seconds, and then instructed to retype the label before proceeding to the next trial. Both targets and distractors were presented in a random order within the following constraints: (i) the pairing of target and distractor varied based on the experimental condition (see §2.4 below for more details on the conditions); (ii) within each training block, each of the 8 meanings appeared three times as a target. The training phase of the experiment consisted of 4 such blocks, each of 24 trials; each block contained the same 24 training trails, with the order of these trials being randomly shuffled.

---

[4] Compliance with the instructions was high – to our knowledge participants did not make use of English or any other language in the current experiment.

2.3.2 Communication Phase

During the communication phase of the experiment, participants took alternating turns as director and matcher:

- **DIRECTOR**: As directors, participants were presented with two images: a target and a distractor. Targets were highlighted with a green border. The director was prompted to type a label that would best communicate the target to the matcher. The label was then sent to the matcher's computer.

- **MATCHER**: Participants were presented with the same two images as the director, with the label provided by the director appearing underneath. The matcher was then prompted to click on the image they thought corresponded to the label provided.

Following each trial, participants were given feedback as to whether or not the matcher had correctly identified the picture described by the director, followed by a display showing the image the director was referring to and the image the matcher selected. Target and distractor pairings were randomly generated within the constraints imposed by the experimental conditions (see §2.4 below), and communication trials were presented in random order. The communication phase consisted of 2 blocks, the length of each block varied depending on the experimental condition (see below).

2.4 Manipulating Context: Mixed, Shape-Same and Shape-Different Conditions

To test the role of context, a simple manipulation was made to the possible combinations of target and distractor images within a single trial during training and communication. This

provides three experimental conditions. For the Shape-Same condition, participants only ever saw pairings of images that shared the same shape, but differed in their idiosyncratic element (see Figure 3A). In the Shape-Different condition, participants were exposed to pairings of images that differed in both their shape and idiosyncratic features (see Figure 3B). Participants in the Mixed condition encountered a mixture of image pairings: some image pairings shared the same shape but differed on their idiosyncratic features, whereas other image pairings differed on both their shape and idiosyncratic features (see Figure 3C).

[Figure Three goes here]

In the Mixed condition, one communication block contained 56 trials, with 24 trials consisting of pairs of images that shared the same basic-level category but differed on subordinate-level features (24 trials exhausting all such possible pairings), whereas the remaining 32 trials differed on both their basic-level category and subordinate-level features (again, 32 trials covering all such possible pairings). To ensure that Shape-Different and Shape-Same conditions were comparable to the Mixed condition in the number of trials, we doubled up the possible combinations of images in the other two conditions, i.e., the Shape-Different condition involved 64 trials (32x2) per communication block and the Shape-Same condition involved 48 trials (24x2) per communication block; participants underwent two such blocks of communication.

2.5   Iteration

The labels produced by a pair of participants in the second block of the communication phase, and their associated target and distractor images, were used to construct the training

language for the next pair of participants: we simply randomly sampled from the communicative output of generation *n* to produce the training language for generation *n+1*, (see Figure 4).

[Figure Four goes here]

The random sampling process was constrained in the following ways. First, for all three conditions, we had a bottleneck on the number of signals that could be passed onto the next generation, i.e., for a single meaning we could only pass on three labels. As such, the number of signals transmitted from the final communication block of a generation stayed consistent between conditions, but the size of the sampling space differed slightly: Mixed (24/56 signals sampled), Shape-Different (24/64 signals sampled), Shape-Same (24/48 signals sampled). Second, in the Mixed condition, the random selection process was additionally constrained so that a given stimuli would appear in at least one shape-same context and one shape-different context, and that that there were an equal number (12) of shape-same and shape-different contexts in total. This meant that, in the Mixed condition, individual stimuli might appear in different ratios of shape-same and shape-different contexts. By contrast, the Shape-Same condition contained all possible pairings of target and distractor in training, and the Shape-Different condition had a subset of all possible contexts (24 out of 32 possible stimuli pairs).

2.6   Dependent Variables and Hypotheses

2.6.1   Measuring Communicative Success

To measure communicative success we simply recorded the number of successful interactions, where the matcher clicked on the target image. Given the differing trial numbers, the maximum success score differs across conditions: Shape-Different (128 points for two blocks of 64 interactions), Mixed (112 points for two blocks of 56 interactions) and Shape-Same (96 points for two blocks of 48 interactions). These maximum scores are converted into proportions to allow visual comparison between the three conditions, but the statistical analyses are conducted on the binary dependent variable.

2.6.2   Measuring Language Types: Difference Scores

In addition to conducting qualitative analyses of the languages that are produced during communication, we used the Normalised Levenshtein edit distance[5] to provide objective measures for WITHIN-CATEGORY DIFFERENCE and BETWEEN-CATEGORY DIFFERENCE. To compute within-category difference for a given block, all labels associated with objects of a given category were compared with one another (i.e., all labels for the 4 blob-shaped images are paired with one another and given a total normalised Levenshtein edit distance, as were all labels for the 4 star-shaped images); the resulting pair of scores (a score for the blob-shaped category and a score for the star-shaped category) were then averaged to obtain a composite within-category difference score. Between-category difference was calculated for a given block by pairing all 4 labels for blobs with all 4 labels for star-shaped images at the same block and calculating average normalised Levenshtein distance.

---

[5] The Normalised Levenshtein edit distance is calculated by taking the minimum number of edits (insertions, deletions, or substitutions of a single character) needed to transform one label into another, and then dividing by the length of the longer label.

These two difference scores provide us with an objective measure of language type. In particular, holistic, systematic and underspecified languages are discriminable on these scores, primarily the within-category difference scores. A holistic language only encodes the idiosyncratic feature of objects in the linguistic system – shape category distinctions are not encoded. As such, we should expect the within-category and between-category differences to be similar. As a systematic language encodes both the shape category and the idiosyncratic element, systematic languages should exhibit smaller within-category difference scores than between-category difference scores, and should also exhibit lower within-category difference scores than holistic languages. For an underspecified language, we expect that only shape category information will be encoded, leading to substantial differences in within-category and between-category difference scores, with within-category scores being close to 0.

2.6.3   Measuring Uncertainty: Conditional Entropy

To further assist in quantifying the language types that emerge, we can calculate the degree of uncertainty in the system, which allows us to quantify the relationship between signals and their associated meaning. First, we need to operationalise two types of uncertainty about signal-meaning pairs. SIGNAL UNCERTAINTY arises from one-to-many pairings of meanings-to-signals (as in cases of synonymy in natural language). Conversely, MEANING UNCERTAINTY arises from one-to-many pairings of signals-to-meanings (as in cases of homonymy and polysemy in natural languages[6]). We predict that the languages in all three conditions will evolve over cultural transmission to lower their signal uncertainty: that is, as a system becomes more conventionalised, it is more likely to only have one signal for each

---

[6] We recognise there is a distinction between *ambiguity*, *vagueness* and *polysemy* in the lexical semantics and cognitive linguistics literature (cf. Tuggy, 1993; Geeraerts, 1993). For the sake of convenience, we use *meaning uncertainty* or *ambiguity* to simply refer to a one-to-many mapping of signals and meanings.

meaning (cf. Reali & Griffiths, 2009). The Mixed and Shape-Same conditions are predicted to evolve toward a one-to-one mapping between signals and meanings (i.e. we should see eight signals for eight meanings in these conditions), leading to low meaning uncertainty. However, the Shape-Different condition is predicted to show higher levels of meaning uncertainty: the prediction is that these chains should involve one-to-many signal-meaning pairs, as an underspecified system leads to the same label being associated with multiple objects which share the relevant feature (here, shape).

To quantify signal uncertainty and meaning uncertainty we measure two aspects of the CONDITIONAL ENTROPY of the system. This gives us a measure of predictability that we can apply to both meaning uncertainty and signal uncertainty. H(M|S) is the expected entropy (i.e. uncertainty) over meanings given a signal, and therefore captures meaning uncertainty,

$$H(M|S) = -\sum_{s \in S} P(s) \sum_{m \in M} P(m|s) \, logP(m|s)$$

where the rightmost sum is simply the entropy over meanings given a particular signal $s \in S$. $P(m|s)$ is the probability that meaning, m is the intended meaning given that signal s has been produced. This entropy is weighted by a distribution $P(s)$ on signals. We can also reverse the position of signals and meanings in this equation to get the conditional entropy of H(S|M), i.e., a measure of signal uncertainty:

$$H(S|M) = -\sum_{m \in M} P(m) \sum_{s \in S} P(s|m) \, logP(s|m).$$

High H(M|S) means that a signal is highly uninformative about the intended meaning (due to the signal having multiple meanings), whilst a high H(S|M) means that a meaning is highly uninformative about the intended signal (due to the meaning having multiple signals).

While these measures capture relevant aspects of the structure of the evolving languages, they do not take context into account, and therefore do not capture the functional adequacy of the system for communication in context. To account for the contextual meaning we incorporate one last measure meaning uncertainty in context, H(M|S, C),

$$H(M|S,C) = -\sum_{s,c\in S,C} P(s,c) \sum_{m\in M} P(m|s,c)\, logP(m|s,c)$$

where the various sums are over signals and meanings GIVEN A CONTEXT. This measure captures the (potential) communicative utility of a system: we predict that the degree of in-context meaning uncertainty will decrease in all three conditions (the languages will be functionally adequate for conveying the correct/intended meaning), whereas meaning uncertainty (disregarding context) will differ across conditions depending on the emerging linguistic systems, as discussed above. As such, we are able to compare these two measures to provide an accurate account of how these types of systems are evolving over time, and whether or not they are adapting to their situational contexts.

2.6.4   Mixed Effects Model Overview

We used R (R Core Team, 2013) and *lme4* (Bates, Maechler & Bolker, 2012) to perform several separate linear mixed effects analyses based on the dependent variables of (a) communicative success, (b) within-category difference scores, (c) between-category difference scores, (d) H(S|M), (e) H(M|S) and (f) H(M|S,C). For our independent variables, we entered condition (Mixed, Shape-Same and Shape-Different), generation and block as fixed effects with interactions. As random effects, we had random intercepts for chain and participant, as well as chain and participant random slopes for generation and block. Each of these models used the Mixed condition as a baseline category. Visual inspection of residual plots did not reveal any noticeable deviations from assumptions of normality or

homoscedasticity. P-values were obtained using a MCMC sampling method (pvals.fnc) provided by the *languageR* package (Baayen, 2008).

2.6.4   Hypotheses

Here we recap and summarise our various hypotheses.

HYPOTHESIS ONE: Participants will increase their communicative success over successive blocks and generations.

HYPOTHESIS TWO: Languages in the Mixed condition will consistently evolve towards systematic category-marking systems.

HYPOTHESIS THREE: Languages in the Shape-Same condition will consistently evolve towards holistic systems.

HYPOTHESIS FOUR: Languages in the Shape-Different condition will consistently evolve towards underspecified systems.

HYPOTHESIS FIVE: The degree of signal uncertainty will decrease across all three conditions over successive blocks and generations.

HYPOTHESIS SIX: The Shape-Different condition is predicted to show higher levels of meaning uncertainty than the Mixed and Shape-Same conditions.

HYPOTHESIS SEVEN: The degree of meaning uncertainty in context will decrease across all three conditions.

3   Results

3.1   Qualitative Results: Languages

This section will provide an overview of a representative selection of languages observed in each of these three conditions. We contrast the initial starting language participants were trained on with very early systems at the start (generation 1, block 1 of communicative interaction) and at the end (generation 1, interaction block 2) of a single generation, as well as systems in the final generation of the chain (generation 3, interaction block 2).

Figure 5 shows an example from chain 1, from the Mixed condition. In generation 1, the labels for each individual referent tend to show some individuation: for instance, MUWUMUWU is only ever associated with one particular blob. However, even at this early stage, we start to see evidence that the labels are patterning systematically according to shape. For instance, the initial syllable MU is consistently associated with blob-shaped referents, and the template H*PA is associated with star-shaped referents. There is also some underspecification: HAPA, for instance, is used with all four stars (albeit at different frequencies). Words lengths also appear to differ systematically between shapes (although this strategy is not repeated in other chains). At the end of the first generation (block 2) a few clear patterns emerge. First, the degree of heterogeneity has decreased in terms of the number of unique words and the number of unique syllables. Second, there is a higher degree of conventionality for each individual referent, as evident in some labels only ever appearing

with one referent (e.g., MUHUMU and HEPA). Lastly, there is less underspecification across star-shaped referents – HAPA is now only associated with two stars. The language of the third generation extends these patterns of increased conventionality: each individual referent has a unique label that distinguishes it from other referents. Furthermore, these labels show systematic relations with one another: three of the blob-shaped images are distinguished from one another through varying the length of (partially) reduplicated syllables (MUWU, MUWUMU and MUWUMUWU). Meanwhile, all of the star-shaped images persist with the basic template of h*pa, and individual referents within this category differ only in the vowel of the first syllable. Finally, there is no underspecification by generation 3: as predicted, the language marks the basic-level category of shape as well as the individuating element. This observation supports our hypothesis that systematic structure will emerge in the Mixed condition, with languages first converging on conventionalised forms for shape followed by the idiosyncratic features.

[Figure Five goes here]

In the first generation of the Shape-Same condition (Figure 6) we see some commonalities with the early stages of the Mixed condition: there are examples of conventionality (e.g., GIGI and ZARA) as well as diversity (e.g., the wide range of labels for the blob with antennae and the star with dots) in the labels used for the individual referents. By time we reach block 2 of the first generation there is almost a completely conventionalised system (in that the participants are aligned on a stable set of labels for each referent). Furthermore, unlike the Mixed condition, this conventionalised system tends to recycle holistic variants instead of introducing systematicity: while there are pockets of systematicity (e.g., KANAKU and NAKAKU), these are circumscribed when compared to the Mixed condition. Interestingly, at

the third generation, the NAPAWE variant has been favoured over the NAKAKU variant, lending additional weight to the notion that the situational context is biasing the system AGAINST systematic structure. These observations provide support for our hypothesis that holistic languages evolve in the Shape-Same condition. However, we should note systematicity is tolerated to a certain extent, as is the case for the blob-shaped images (KAPA and KAPAPA and GUGU and GIGI).

[Figure Six goes here]

For the Shape-Different condition (Figure 7), we see that there is a high level of heterogeneity in both the labels used between and within the referents. There is, however, some clustering of syllable types (e.g., NO, GO, NI etc) and combinatorial patterns (e.g., PUGO, GOGO, PUMA) according to the basic-level category of shape. Interestingly, this diversity persists in the first generation (block 2), with less conventionality than that found in the Mixed and Shape-Same conditions. Still, there is an increase in conventional patterns, with forms becoming more predictable over time in both the number of syllables and the way in which they are arranged (e.g., ME and HE tend to disproportionately occur in the initial syllable position). The most noticeable difference between generation 1 and generation 3 is the collapse towards underspecification: we see high frequency forms for all blob-shaped referents (e.g., PUGU) and all star-shaped referents (e.g., HEHA). In addition to this loss of variation at the word level, variation also decreases at the syllable level (e.g., there are only four syllables for blob-shaped images: PU, PO, GU, and GO). The emergence of underspecified languages supports our hypothesis that languages in Shape-Same condition will evolve to abstract across the meaning dimension of shape.

[Figure Seven goes here]

It is important to note that all three conditions started off with a language that consists of randomly generated pairings of labels and meanings. Although the individual pairings differ between conditions, they do share an important structural characteristics: all initial languages have high levels of synonymy (three labels for each meaning). A consistent pattern shared across all three conditions is a shift from this system with many-to-one signal-meaning mappings to systems where we observe one-to-one and one-to-many mappings.

3.2   Communicative Success

Communicative Success scores tended to follow a similar trajectory in all three conditions (see Figure 8). Over successive blocks we observe a clear increase in the overall communicative success rate, leading to near-perfect communication by the end of generation 3. Analysis of the logistic mixed effects model revealed a significant main effect of Generation ($\beta = 1.13$, SE $= 0.19$, z $= 6.646$, p $< .001$) and Block ($\beta = 1.03$, SE $= 0.30$, z $= 3.399$, p $< .001$), but no effect of Condition and no other significant interactions (p $>.074$).

[Figure Eight goes here]

These results show that, in all conditions, the languages are becoming increasingly effective at achieving communicative success through (a) repeated interactions between individual participant pairs and (b) across successive generations of participant pairs.

3.3   Difference Scores

Table 1 shows the idealised and observed (in the second block of generation 3) values for the within- and between-category difference measures. Figure 9 shows how these measures evolve over time. Our hypothesis that languages in the Mixed condition should evolve systematic category-marking and should therefore produce a within-category difference score of around 0.5 (characteristic of a system in which signals tend to be composed of a general category-marker and an individuating element) and a between-category difference score of 1 (distinctive labels used across categories). For the Shape-Same condition, we predicted the emergence of holistic languages, where each object is associated with a unique and distinctive label: this is characterised by high within- and between-category differences. As can be seen from Table 1, these predictions were borne out. For the Shape-Different condition, we predicted the emergence of systems that underspecified, using a single label for all objects sharing a shape, which would correspond to 0 within-category difference and a high between-category difference: as can be seen from the table, while this prediction was partially supported (within-category difference is lower than between-category difference), the within-category difference in this condition remains high – this is due to the slower conventionalisation seen in this condition, as highlighted in the qualitative analysis above (see also measures of signal uncertainty below).

[Table 1 goes here]

[Figure Nine goes here]

Analysis of the mixed-effects model for Within-Category difference showed a significant effect of Generation (β = -0.07, SE = 0.02, t(84) = -2.823, p <.001), and a significant main effect of Shape-Same condition (β = 0.20, SE = 0.04, t(84) = 5.043, p <.001). There was one

significant interaction for Shape-Same condition x Generation (β = 0.07, SE = 0.03, t(84) = 2.266, p =.017). All other main effects and associated interactions were non-significant (p >.061). These results partially support our predictions: within-category difference remains high in the Same-Shape condition, reflecting the development of labels which individuate within categories, and decreases in the other conditions; however, the Within-Category differences remain surprisingly high in the Shape-Different condition, where we predicted the emergence of a fully underspecified system, associated with Within-Category difference of 0.

Analysis of the model for Between-Category difference showed that only the main effect of Generation was significant (β = 0.04, SE = 0.02, t(84) = 2.46, p <.001), supporting the contention that Between-Category labels become increasingly distinct from one another over generations. All other main effects and associated interactions were non-significant (p >.139).

3.4   Conditional Entropy

3.4.1   Signal Uncertainty H(S|M)

For the conditional entropy of signals given meanings, H(S|M), we observe a general decrease across all three conditions (see fig. 10). However, the decline in entropy for the Shape-Different condition appears to be less pronounced than that of the Mixed and Shape-Same conditions: as discussed above, within-category variation persists unexpectedly in this condition. For H(S|M) the mixed-effects model contained significant results for the main effects of Generation (β = -0.61, SE = 0.13, t(72) = -4.561, p <.001), Block (β = -0.38, SE = 0.09, t(72) = -4.366, p <.03) and Shape-Different condition (β = 0.62, SE = 0.31, t(72) = 2.011, p <.009). There were no other significant main effects or interactions (p >.259).

[Figure Ten goes here]

3.4.2   *Meaning Uncertainty H(M|S)*

Figure 11 plots the conditional entropy of meanings given signals, H(M|S), against the number of blocks. As predicted, there is a clear difference between the conditions, with the Shape-Different condition showing a general increase in entropy in contrast to the Mixed and Shape-Same conditions, corresponding to the development of underspecified labels. For H(M|S) the mixed-effects model contained significant results for the main effect of the Shape-Different condition (β= 0.41, SE = 0.10, t(72) = 4.053, p<.001). There was also a significant Shape-Different condition x Block interaction (β = 0.32, SE = 0.09, t(72) = 3.424, p <.001). There were no other significant main effects or interactions (p >.265).

[Figure Eleven goes here]

3.4.3   *Meaning Uncertainty of signals in context H(M|S,C)*

The conditional entropy of meanings given signals in context, H(M|S,C), is shown in figure 12. In all three conditions we observe a decrease in entropy over time, with each of the conditions showing strikingly similar trajectories of change: as indicated by the communicative accuracy scores, the languages in all conditions evolve towards allowing optimal communication in context. For H(M|S,C) the mixed-effects model contained significant results for the main effects of Generation (β = -0.08, SE = 0.03, t(72) = -3.300, p <.001) and Block (β = -0.07, SE = 0.01, t(72) = -5.927, p <.001). There were no other significant main effects or interactions (p >.078).

[Figure Twelve goes here]

4 Discussion

Our findings support the general hypothesis that language structure adapts to the situational contexts in which it is learned and used. As we outlined in the introduction, some meaning is encoded and some meaning is inferred, with interactional short-term strategies of conveying the intended meaning feeding back into long-term, system-wide changes. In our experiment, languages gradually evolved to encode information relevant to the task of achieving communicative success in context, with different language systems evolving in each experimental condition. In the Shape-Same condition, where the dimension of shape was always the same for stimuli pairings, holistic systems of communication emerged, whilst in the Shape-Different condition, where the dimension of shape was always different for stimuli pairings, the system generalised and became underspecified (although unexpectedly variable: see discussion below). For the Mixed condition, which featured both Shape-Same and Shape-Different contexts, the systems that emerged were systematically structured: that is, both shape category and individual identity were encoded in the linguistic signal. These divergent systems arise given a very simple meaning space, through slight manipulations to the situational context.

Despite these inherent differences between the languages that emerged, all of the conditions showed: (a) an increased level of communicative success and (b) a reduction in in-context meaning uncertainty, $H(M|S,C)$. This observation suggests each condition produces languages that are functionally adequate for the task of achieving communicative success in

context. The fact that different systems evolve for conveying the same set of meanings is important for how we view the role of context. Our explanation rests on the premise that languages are adapting to their niche, which in this case comprises the situational context, to become optimally structured.

Underspecified systems emerge in the Shape-Different condition because "when context is informative, any good communication system will leave out information already in the context" (Piantadosi, Tily & Gibson, 2012: 284). This lends weight to studies showing that participants are making use of pragmatic reasoning to convey information at the least cost given common knowledge and the task at hand (Frank & Goodman, 2012). These underspecified systems could be construed as being highly ambiguous when taken out of their communicative context. However, when we take into account the context in which the signals were used (as measured by the $H(M|S,C)$) then the apparent ambiguity is not counter-functional: that is, the system is perfectly adequate for achieving communicative success. When examined out of context, adapted communication systems can give the appearance of ambiguity, as Miller (1951: 111-2) noted: "Why do people tolerate such ambiguity? The answer is that they do not. There is nothing ambiguous about 'take' as it is used in everyday speech. The ambiguity appears only when we, quite arbitrarily, call isolated words the unit of meaning."

While the amount of synonymy (as measured by $H(S|M)$) decreased over time across all conditions, the Shape-Different condition appeared to tolerate a higher level of synonymy than the other two conditions. One possible explanation is the way in which participants viewed the task. An initially diverse input could be construed as priming the participants to reproduce a diverse output. If the labels are easy enough to learn and reproduce, and they

achieve the goal of successfully allowing the matcher to choose the correct image, then this variation may be tolerated for longer. This also partly explains why the Shape-Different condition deviates from its predicted within-category difference score: labels are not conventionally associated with any one particular meaning within a category. For instance, as discussed in the qualitative analysis (see Fig. 7), PUGU and POGO (which are quite distinct, with a normalised Levenshtein edit distance of 0.5) are not conventionally associated with any particular blob; instead, they pattern synonymously, with the two labels being optional forms for *any* blob-shaped image. This reflects a limitation of the difference measurement to distinguish between systematic languages and this kind of synonymy. However, these languages do have distinct profiles, as evidenced by the various entropy measurements.

It is also worth noting that not all chains in the Shape-Different condition converged on an underspecified system, with chain 11 evolving a holistic-like system. This mismatch with our predictions is perhaps due to the Shape-Different condition having more optionality provided by the situational context: that is, any of three hypothesised systems (Underspecified, Holistic, Systematic) are expressively adequate for conveying the intended meaning, although these systems differ in their parsimony in terms of memory and learning demands. This increases the probability that we will see more variation in the types of systems that evolve in the Shape-Different condition. Whereas underspecified and, to a lesser extent, systematic category-marking languages are communicatively sub-optimal in the Shape-Same condition, the Shape-Different condition does not share such restrictions. A similar story applies when comparing the Mixed and Shape-Different conditions: neither holistic nor systematic category-marking languages are disfavoured for either condition, but an underspecified system would be problematic in the Mixed condition (as 43% of the contexts have images

that share the same shape). Chain 11 thus serves as an important reminder of lineage-specificity, and how the historical properties of a particular system can bias future states.

For the Shape-Same condition, the chains consistently converge on holistic systems: that is, each individual stimulus has a unique label, with these labels being relatively distinct from one another. The decrease in H(S|M) and H(M|S) shows that the system is converging towards a one-to-one mapping of forms and meanings, whereas the high within-category difference scores show these signals are highly distinct from one another, and indeed more distinctive than those found in the other two conditions. Our rationale for the emergence of holistic systems in the Shape-Same condition is similar to that of the Shape-Different condition: where the situational context is informative, information will be left out of the linguistic system. In this instance, the context was informative through virtue of having the pairs of stimuli always sharing the same shape. This explains why systematicity is minimised in the Shape-Same condition: the linguistic system does not need to conventionally encode shape into the signal because context makes it irrelevant in discriminating between meanings. Instead, these languages specialise and become holistic, allowing them to meet the participants' communicative needs in context.

Even though the languages which emerge in the Shape-Same condition do reliably differ from those that evolve in the Mixed condition, through being more holistic, there is some evidence of systematicity in these chains. In chain 6, for instance, a language evolved in which two of the blob-shaped stimuli share similar labels (KAPA and KAPAPA) as do the other two blob-shaped stimuli (GUGU and GIGI). These pockets of correlations between word forms suggest a certain degree of systematicity is tolerated – albeit not to the same extent as that found in the Mixed and Shape-Different conditions. One explanation for this finding is that

the situational context and communication are not the only factors shaping the system, with learnability pressures also acting on the structure of language (Kirby, Cornish & Smith, 2008).

Only in the Mixed condition do we consistently observe the emergence of systematic category-marking languages. The first line of evidence is that the observed within-category difference score lines up with our expected score (see Figure 11): this suggests part of the label is specifying shape and the other part is specifying the individuating component. While, as noted above, a difference score of approximately 0.5 is not necessarily indicative of systematic language structure, the H(S|M) and H(M|S) scores show that, by generation 3, the languages in the Mixed condition have low conditional entropy, showing that the form-meaning pairs embody one-to-one mappings.

A holistic language would be just as successful as conveying the correct meaning as a systematic language in the Mixed condition. So why do we see the emergence of systematic instead of holistic languages? Part of the reason rests on how these languages evolve in the early stages of their emergence: participants quickly establish a conventionalised specification of shape, before arriving upon conventionalised forms that encode the individuating elements. As a strategy, specifying shape information only requires participants to align on two signals, one that specifies star-shaped objects and one that specifies blob-shaped objects, which would allow them to successfully communicate on 57% of trials (those where discrimination only requires that shape information is conventionally encoded).

We can view this strategy as a negotiated exploration of the specification space during interaction, giving rise to a two-stage process: (i) THE CONVENTIONALISATION OF

CATEGORY-MARKING FOR SHAPE; (ii) THE CONVENTIONALISATION OF INDIVIDUATING ELEMENTS. Supporting this contention of a two-stage process is the main effect of Generation for both the within-category difference scores and the conditional entropy of H(S|M): even though the within-category difference scores suggest systematic category-marking emerges by the end of generation one, the H(S|M) entropy is much higher in this initial generation than it is at later generations. The decrease in H(S|M) reflects the conventionalisation of individuating elements in the linguistic system - that is, there is less synonymy in later generations.

Another striking finding in the Mixed condition was the rate at which systematic category-marking emerged, within a single generation of participants. Part of the explanation could be in how the manipulation of context exerts a strong constraint for participants to quickly converge on conventional markers for shape. There are several reasons why the rapid evolution seen in this experiment might prove to be an exception, rather than a general tendency. First of all, there are only two possible dimensions that the language may encode: the basic-level category and the subordinate idiosyncratic component. There are also differences between the initial generation and successive generations (as mentioned above): namely, later generations show greater degrees of conventionalisation in their label usage.

If languages are adapting to their contextual niche, then what are the implications for the learning bias and historical contingency accounts? Even though our results are broadly consistent with the ecologically sensitive account, there is also evidence consistent with the learning bias (e.g., pockets of systematicity in the Shape-Same condition and the overall reduction of synonymy across all conditions) and historical contingency (e.g., the emergence of a holistic language in chain 11 of the Shape-Different condition) accounts. It is likely that

all these theoretical perspectives hold true to some extent, with the role of context being mediated by partially-competing motivations of prior learning biases and historical contingency. Such notions reflect the converging evidence that languages, and the way in which they are organised, "are better explained as stable engineering solutions satisfying multiple design constraints, reflecting both cultural-historical factors and the constraints of human cognition" (Evans & Levinson, 2009: 429).

5   Conclusion

We set out to investigate the role of situational context in the emergence of different types of linguistic systems that evolve through iterated learning. By manipulating the ways in which stimuli were paired with one another, we showed that situational context is an important factor is determining what is and is not encoded in the linguistic system. Our results offer a potential insight into how the situational context can bias the cultural evolution of language. The type and predictability of the situational contexts relate to how language users will employ certain communicative strategies for conveying the intended meaning, with the resulting language systems reflecting the contextual constraints in which they evolved.

One of the major findings in our experiment is that the types of linguistic systems that evolve are highly predictable based on their contextual constraints during communication. This interplay between short-term linguistic strategies for resolving communicative interactions, and the implication for language systems through long-term patterns of change, speaks to real-world processes such as grammaticalisation: the types of change we observe in languages show predictable patterns, as evident in the unidirectionality hypothesis (cf. Hopper & Traugott, 2003), but importantly these changes show how contextual constraints on

the moment-to-moment communicative strategies deployed can have widespread ramifications on whole linguistic systems (Steels, 2012). Natural languages are subject to a larger and more diverse range of contexts, with a key future question being the extent to which our experimental results are generalisable to patterns observed in natural language systems.

6   References

Altmann, E.G., Pierrehumbert, J.B., & Motter, A.E. (2011). Niche as a Determinant of Word Fate in Online Groups. *PLoS ONE*, 6(5): e19009. doi:10.1371/journal.pone.0019009.

Ay, N., Flack, J.C. & Krakauer, D.C. (2007). Robustness and complexity co-constructed in multimodal signaling networks. *Philosophical Transactions of the Royal Society B*, 362: 441-447.

Baayen, R.H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bach, K. (2012). Context Dependence (such as it is). In *The Continuum Companion to the Philosophy of Language*, M. Garcia-Carpintero and M. Kolbel, eds.

Bates, D.M., Maechler, M. & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M.H., Croft, W., et al. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, 59: 1–26.

Beuls, K., & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLoS ONE* 8(3): e58960. doi: 10.1371/journal.pone.0058960.

Bleys, J. & Steels, L. (2009). Linguistic Selection of Language Strategies - A Case Study for Colour. *ECAL*, 2: 150-157.

Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review*, 101: 327-341.

Brighton, H., Kirby, S., & Smith, K. (2005). Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In Tallerman, M., editor, *Language Origins: Perspectives on Evolution*, chapter 13. Oxford: Oxford University Press.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Caldwell, C.A. & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocities. *PLoS ONE*, 7(8): e43807. doi: 10.1371/journal.pone.0043807.

Chafe, W. (1976). Giveness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. In *Subject and Topic*, ed. Charles N. Li, 25-56. New York: Academic Press.

Christiansen, M.H. & Chater, N.H. (2008). Language as shaped by the brain. *Behavioral and Brain Science*, 31(5): 489-508.

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cornish, H., Tamariz, M. & Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: what experiments can tell us. Special issue on Language as a Complex Adaptive System. *Language Learning*, 59(s1): 187-205.

Culbertson J, & Adger D. (2014). Language learners privilege structured meaning over surface frequency. *PNAS*, 111(16): 5842-5847.

Culbertson, J., Smolensky, P., & Wilson, C. (2013). Cognitive biases, linguistic universals, and constraint-based grammar learning. *Topics in Cognitive Science*, 5(3):392–424.

Croft, W. (2000). *Explaining language change: An evolutionary approach*. Harlow, Essex: Longman.

Dingemanse, M., Torreira, F., & Enfield, N.J. (2013). Is "Huh?" a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items. *PLoS ONE*, 8(11): e78273. doi:10.1371/journal.pone.0078273.

Du Bois, J. W. (1987). The Discourse Basis of Ergativity. *Language*, 63(4): 805-855.

Dunn, M., Greenhill, S.J., Levinson, S.C., & Gray, R.D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473: 79-82.

Evans, N. & Levinson, S.C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(5): 429-492.

Evans, V. & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Evans, V. (2005). The meaning of time: Polysemy, the lexicon and conceptual structure. *Journal of Linguistics*, 41: 33-75.

Fay, N., & Ellison, T.M. (2013). The Cultural Evolution of Human Communication Systems in Different Sized Populations: Usability Trumps Learnability. *PLoS ONE*, 8(8): e71781. doi:10.1371/journal.pone.0071781.

Fedzechkina, M., Jaeger, T.F., & Newport, E.L. (2012). Language learners restructure their input to facilitate efficient communication. *PNAS*, 109: 17897-17902.

Ferreira, V., Slevc, L., & Rogers, E. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, 96(3): 263-284.

Fery, C. & Krifka, M. (2008). Information structure. Notional distinctions, ways of expression. *Unity and diversity of languages*, ed. Piet van Sterkenburg. Amsterdam: John Benjamins, 123-136.

Frank, M.C., and Goodman, N.D. (2012). Predicting Pragmatic Reasoning in Language Games. Science, 336: 998. doi: 10.1126/science.1218633.

Galantucci, B., Garrod, S., and Roberts, G. (2012). Experimental Semiotics. *Language and Linguistics Compass*, 6: 477–493. doi: 10.1002/lnc.

Galantucci, B., Kroos, C., and Rhodes, T. (2010). The effects of rapidity of fading on communication systems. *Interaction Studies,* 11: 100–111.

Garrod, S., & Galantucci, B. (2011). Experimental Semiotics: A Review. *Frontiers in Human Neuroscience*, 168 (6).

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science,* 31: 961–987.

Geeraerts, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics*, 4(3):223-272.

Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62: 451-82.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66: 377-388.

Griffiths, T.L., & Kalish, M. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31: 441-480.

Heine, B., Claudi, U., & Hünnemeyer, F. (1991). *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press.

Hoefler, S.H. (2009). *Modelling the role of pragmatic plasticity in the evolution of linguistic communication*. PhD Thesis, University of Edinburgh.

Hopper, P. J. & Traugott, E. C. (2003). *Grammaticalization*, Cambridge Textbooks in Linguistics: Cambridge University Press.

Kay, P. (1977). Language evolution and speech style. In: Blount, B.G. & Sanches, M. (Eds.), *Sociocultural dimensions of language change*. Academic Press: New York, pp. 21-33.

Kirby, S. (1999). *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford: Oxford University Press.

Kirby, S. (2012). Language is an Adaptive System. The Role of Cultural Evolution in the Origins of Structure. In M. Tallerman & K. R. Gibson (Eds.), *The Oxford Handbook of Language Evolution*: 589-604. Oxford: Oxford University Press.

Kirby, S., Cornish, S. & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, 105: 10681-10686.

Kirby, S. & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language*. London: Springer Verlag, pp 121-148.

Kamide, Y., Altmann, G., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1): 133-156.

Levinson, S. (2000). *Presumptive meanings: The theory of generalised conversational implicature*. MIT Press.

Lass, R. (1997). *Historical linguistics and language change*. Cambridge University Press: Cambridge.

Lewis, M., Sugarman, E., & Frank, M. C. (2014). The structure of the lexicon reflects principles of communication. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society.*

Lupyan, G. & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5(1): e8559. doi: 10.1371/journal.pone.0008559.

Lyons, J. (1977). *Semantics*. Cambridge University Press: Cambridge.

Miller, G.A. (1951). *Language and Communication*. McGraw-Hill Book Company: New York.

Perfors, A. & Navarro, D.J. (2014). Language evolution can be shaped by the structure of the world. *Cognitive Science*, 38(4): 775-93.

Piantadosi, S.T., Tily, H. & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122: 280-291.

Pinker, S. & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13: 707-784.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thrope, K. (2010). The effects of Feature-Label-Order and their implications for symbolic learning. *Cognitive Science*, 34: 909-957.

Reali, F. & Griffiths, T.L. (2009). The evolution of linguistic frequency distributions: Relating regularisation to inductive biases through iterated learning. *Cognition*, 111: 317-328.

Scott-Phillips, T.C., Kirby, S., & Ritchie, G.R.S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(2): 226-33.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, 14(9), 411-417.

Sedivy, J. C. (2005). Evaluating explanations for referential context effects: Evidence for Gricean Mechanisms in Online Language Interpretation. In J.C. Trueswell & M.K. Tanenhaus (eds.): *Approaches to studying world-situated language use: Bridging the language as product and language as action traditions*, MIT Press, Cambridge, MA.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *PNAS*, 104: 7361–7366.

Silvey, C., Kirby, S. & Smith, K. (2014). Word meanings evolve to selectively preserve relevant distinctions over cultural transmission. *Cognitive Science*. doi: 10.1111/cogs.12150.

Smith, K., Tamariz, M. & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1348-1353). Austin, TX: Cognitive Science Society.

Smith, K. & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116: 444-449.

Sperber, D. & Wilson, D. (1995). *Relevance: communication and cognition*, Second Edition, Oxford/Cambridge: Blackwell Publishers.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Sciences*, 7(7): 308-312. doi: 10.1016/S1364-6613(03)00129-3.

Steels, L. (2012). Self-organization and Selection in Cultural Language Evolution. In: Luc Steels, ed., *Experiments in Cultural Language Evolution*. John Benjamins: Amsterdam.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, Mass.: MIT Press.

Thiesen-White, C., Kirby, S. & Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. In: *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pp. 956-961.

Traugott, E. & Konig, E. (1991). The semantics-pragmatics of grammaticalization revisited. In *Approaches to Grammaticalization*, eds., Elizabeth Traugott and Bernd Heine: 189-218.

Traugott, E. & Trousdale, G. (2013). *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Tuggy, D. (1993). Ambiguity, polysemy and vagueness. *Cognitive Linguistics*, 4: 273-291.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, 4(4): 357-380.

Wedgwood, D. (2007). Shared assumptions: Semantic minimalism and relevance theory. *Journal of Linguistics*, 43: 647-681.

Wray, A. & Grace, G.W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117: 543-578.

Zipf, G. (1949). *Human behavior and the principle of least effort*. New York: Addison-Wesley.

Table One

Condition	Within-Category Difference		Between-Category Difference	
	Idealised	Observed	Idealised	Observed
Shape-Same	1	0.74 (SD = .05)	1	0.80 (SD = .07)
Mixed	0.5	0.47 (SD = .07)	1	0.88 (SD = .05)
Shape-Different	0	0.62 (SD = .10)	1	0.90 (SD = .08)

Table 1. The idealised (left-hand columns) and observed (right-hand columns) scores for within-category differences and between-category differences. Numbers in brackets indicate the bootstrapped standard deviation.

Figure One
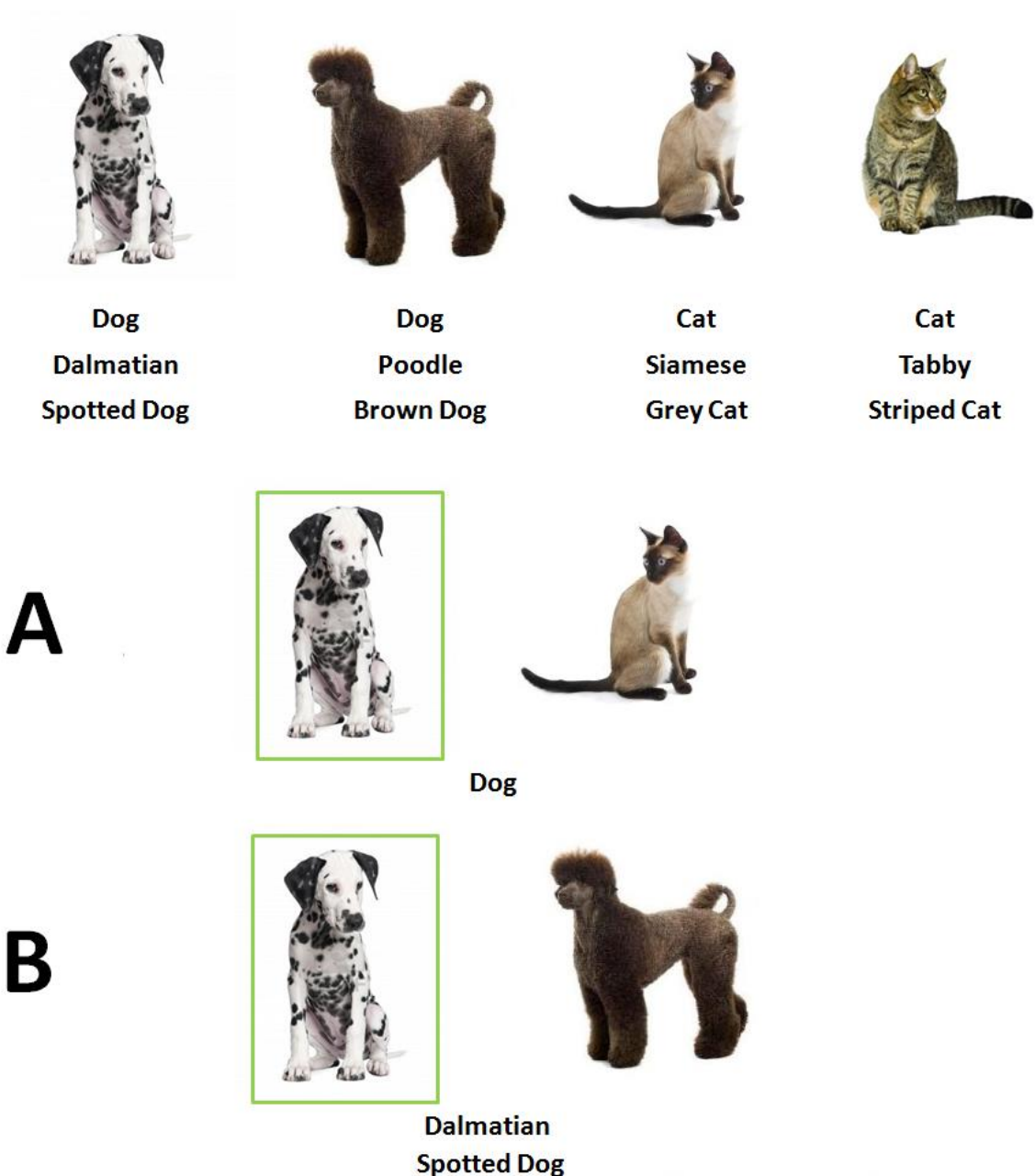
Figure 1. Language strategies and example contexts. The green coloured boxes correspond to the intended referent. As we can see, DOG is a viable strategy for conveying the intended meaning in context A, but we need to either use a more specific label (DALMATIAN) or provide additional referential information alongside the generalised form (SPOTTED DOG) to convey the intended meaning in context B.

Figure Two
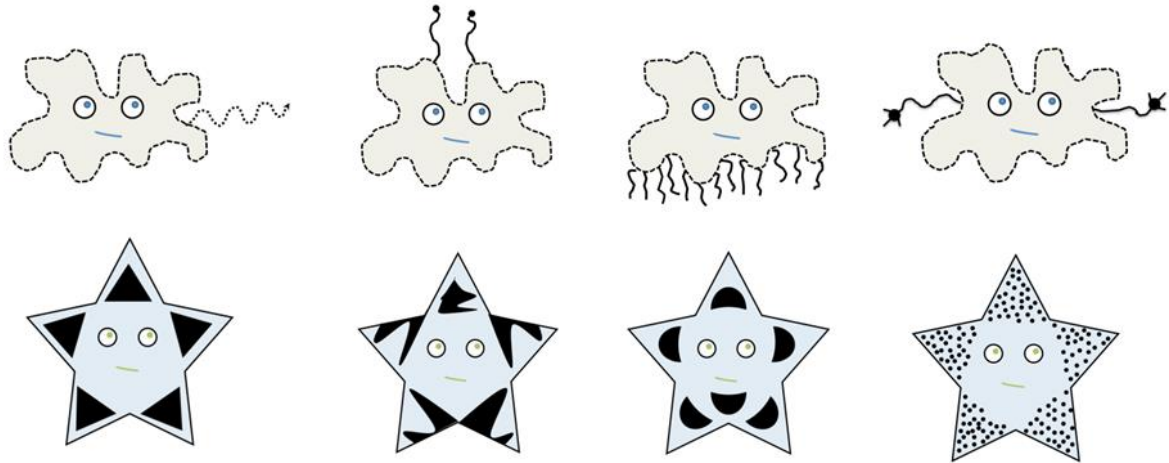
Figure 2. All eight meanings for the image stimulus set used in this experiment. Note that each individual image is comprised of two components: a basic-level of shape (star or blob) and a subordinate-level (a unique idiosyncratic feature).

Figure Three

A
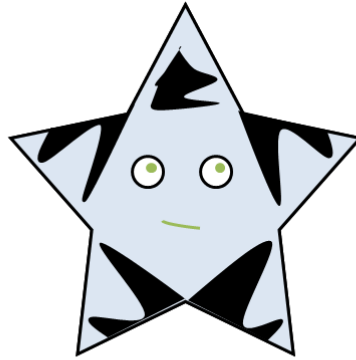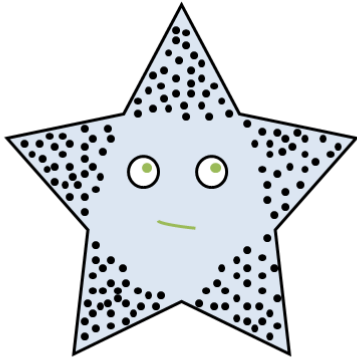
B
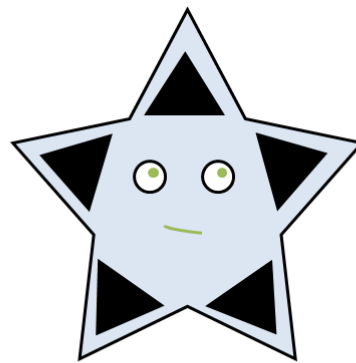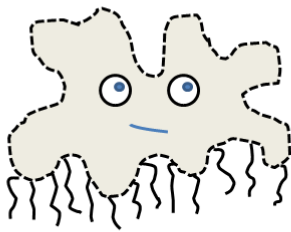
C
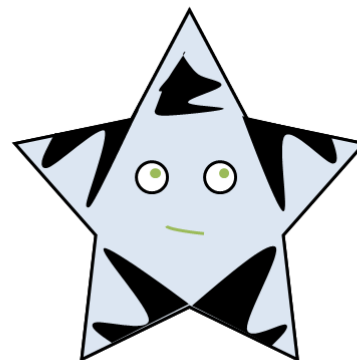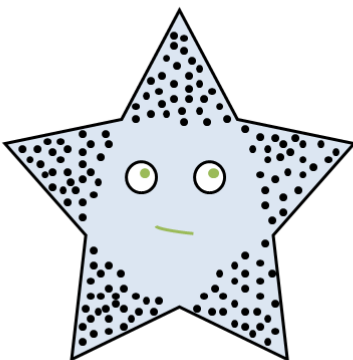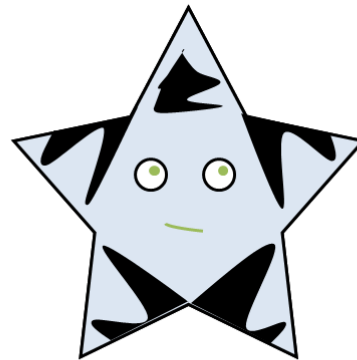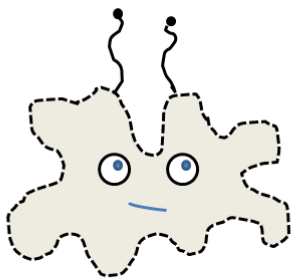
Figure 3. **3A** is an example of a pairing in the Shape-Same condition: here, participants only ever observe pairings that share the same shape. **3B** is an example of a pairing found in the Shape-Different condition; the two stimuli always differ in shape. **3C** shows an example of the pairings used in the Mixed condition: here, we get a mixture of stimuli that in some contexts differ in shape and in other contexts share the same shape.

Figure Four
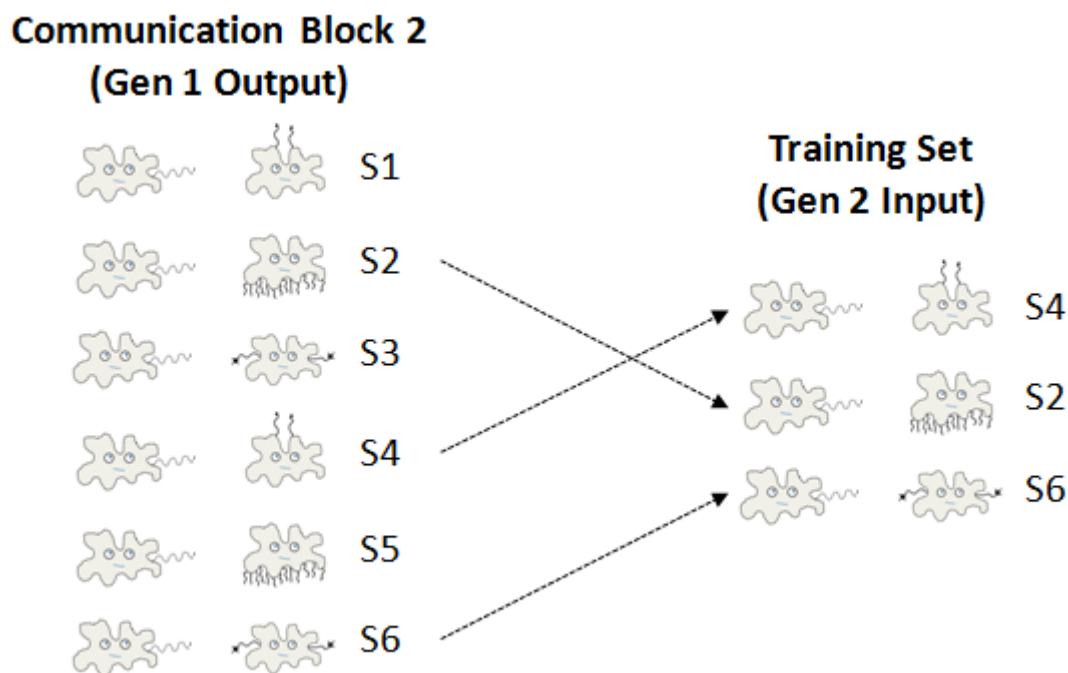
Figure 4. An example of the random selection process employed for a single meaning in the Shape-Same condition. Here, one target meaning is associated with 6 (possibly unique) signals during communicative testing. However, only three trials are required to construct a training block for the next generation: in order to generate this training block, we sample randomly from the appropriate contexts.

Figure Five

Figure 5. A table showing the initial training language and all of the signal-meaning pairs produced at generation 1 (communication block 1), generation 1 (communication block 2) and generation 3 (communication block 2) in chain 1 (Mixed condition). Each meaning appears with a collection of labels beneath it: this constitutes the combined output of a pair of participants in a particular generation.

Figure Six

Input	gimu / lagihu / kapa	moka / gipawe / wepa	gila / gigi / kala	mohupapa / pakamo / wegu	nagimuhu / mogikamu / hukahu	wemunahu / hupa / kawena	kalaka / napawe / nahu	kuhunamu / nagimu / kamuwemo	**Generation One** *Training data*

Output	gugu	moka	gigi	wepapa	hukahu	zara	zala	bobu	**Generation One** *Block One*
	gugu	gaga	gigi	wepapa	hukahu	zara	kalaka	pada	
	kapa	moka	gigi	wepapa	hakahu	zara	napawe	pada	
	kapa	gugu	gigi	kapa	hukawe	zara	zala	bobu	
	trala	gugu	gigi	wepapa	hukawe	zara	kalaka	bobu	
	wepa	gugu	gigi	kapapa	hokuwe	zara	kalaka	bobu	

Output	kapa	gaga	gigi	kapapa	kuluwe	zara	napawe	bobu	**Generation One** *Block Two*
	kapa	gugu	gigi	kapapa	kuluwe	zara	napawe	bobu	
	kapa	gugu	gigi	kapapa	kuluwe	zara	napawe	bobu	
	kapa	gugu	gigi	kapapa	kanaku	zara	napawe	bobu	
	kepa	gugu	gigi	kapapa	kanaku	zara	nakaku	bobu	
	kapa	gugu	gigi	kapapa	kanaku	zara	nakaku	bobu	

Output	kapa	gugu	gigi	gugu	kanaku	zara	napawe	bobu	**Generation Three** *Block Two*
	kapa	gugu	gigi	kapapa	kanaku	zara	napawe	bobu	
	kapa	gugu	gigi	kapapa	kanaku	zara	napawe	bobu	
	kapa	gugu	gigi	kapapa	kanaku	zara	napawe	bobu	
	kapa	gugu	gigi	kapapa	kanaku	zara	napawe	bobu	
	kapa	gugu	gigi	kapapa	kanaku	zara	napawe	bobu	

Figure 6. A table showing the initial training language and all of the signal-meaning pairs produced at generation 1 (communication block 1), generation 1 (communication block 2) and generation 3 (communication block 2) in chain 6 (Shape-Same condition).

Figure Seven

Input	gugogo pupuheno pugono	higo nopuma gomahehi	henoni puhi puhanino	hemahiha puhigohi gugonini	hehima heha guhigono	nipuni hegoni nohihama	manohapu mahi gogohani	hamahi nogoma hagonihi	**Generation One** *Training data*
Output	nogogo nopuma pugoni nogogi pugoni pugoni pugoni nonogini	nopuma nogogi nepuma nopuma puma puma puma nugoni	pugoni pugoni pugoni nepuma pupunogi nopuma gogononi pugoni	gogonin gogonin nopuma puhani pugoni pupunogini pumagogi nugoni	hehami heha heha nehema nehema heramah nohema nemahena	manohi minaha nemaha hamahi nemahihi hamahi hamahi hamahi	hemaha hemaha hemaha hemaha henema nehema memeha memeha	manoha hamahi memaha nehaha hamahi hemeha hamahi hamahi	**Generation One** *Block One*
Output	nugino nugogo nogogo nupogo gogoni gogoni nugoni pugogo	puma puma puma puma nugoni pupino monogo pum	pupunoni pupunino pupuno pugoni pugini pupuno pogo punipu	punino pupino nopugo pupon nugopu punogo nopuma monogo	nehema nehema hemena hemaha namaha mahima mahima heha	hamahi nehema nehema nehema hamahi hemena hemena he	mehama mehama mehama mehama mehima mahima heha hehama	maneha nehaha hamahi hehema meneha hemena harama hemana	**Generation One** *Block Two*
Output	pugu pugu pogo pogo pugogo pugogo pugu pugu	pugu pugo pugu pugo pogo pugo pugu pugogo	pugu pugo pugu pugo pogo pogo pugu pugu	pugu pugogo pogo pugo pogo pogo pugu pugu	heha heha heha heha heha heha heha heha	heha heha mehima meheha heha heha heha meheha	heha heha heha heha heha heha heha mehima	heha heha heha heha heha heha hemaha mehima	**Generation Three** *Block Two*

Figure 7. A table showing the initial training language and all of the signal-meaning pairs produced at generation 1 (communication block 1), generation 1 (communication block 2) and generation 3 (communication block 2) in chain 12 (Shape-Different condition). Highlighted labels show underspecification.
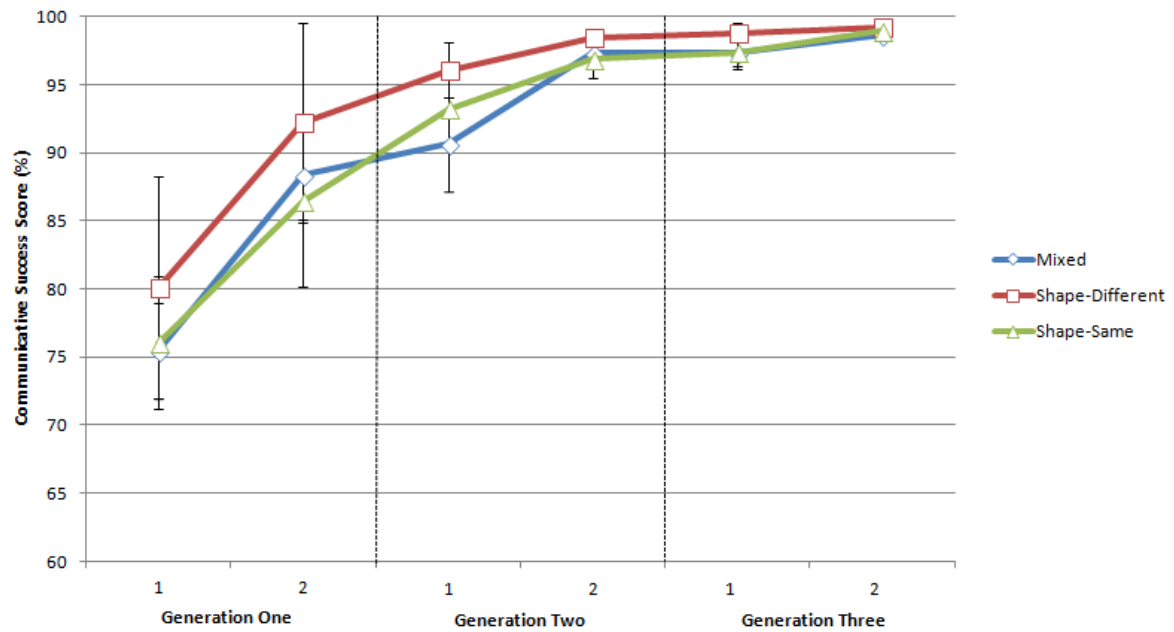
Figure Eight

Figure 8. Average communicative success scores by generation (1-3), communication block, and condition. The vertical dotted lines represent the start of the next generation. Error bars represent the 95% confidence intervals.
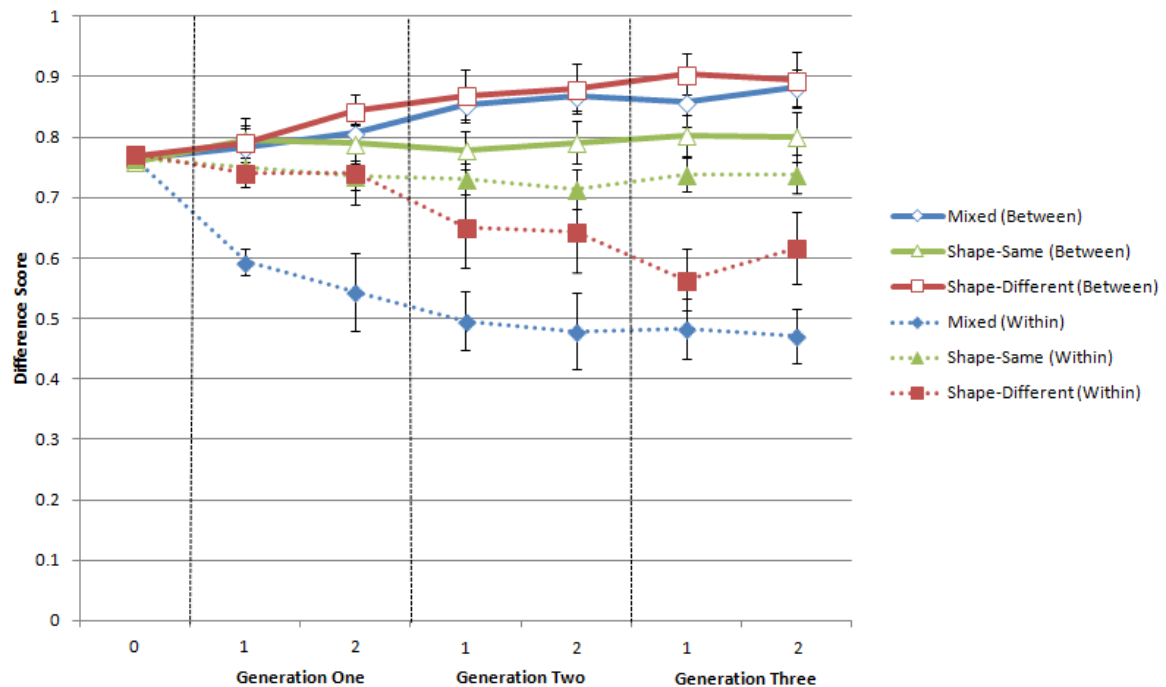
Figure Nine

Figure 9. Between-category (solid lines) and within-category (dotted lines) difference scores (measured by the average normalised Levenshtein edit distance) over successive

communication blocks for the Mixed (blue lines), Shape-Same (green lines) and Shape-Different conditions (red lines). Generation 0 gives values for the initial random language. Error bars indicate 95% confidence intervals.
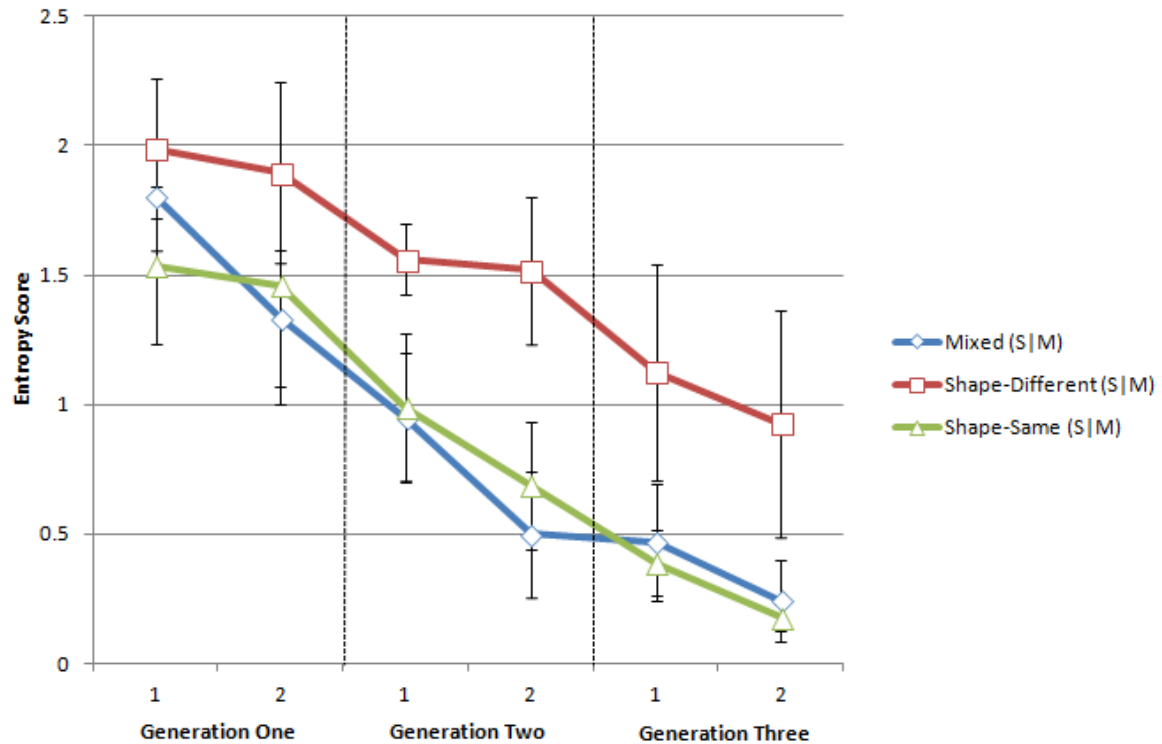
Figure Ten

Figure 10. Degree of signal uncertainty, measured as H(S|M), against Generation and Block. Higher entropy scores indicate a higher degree of signal uncertainty. The error bars indicate the 95% confidence intervals.
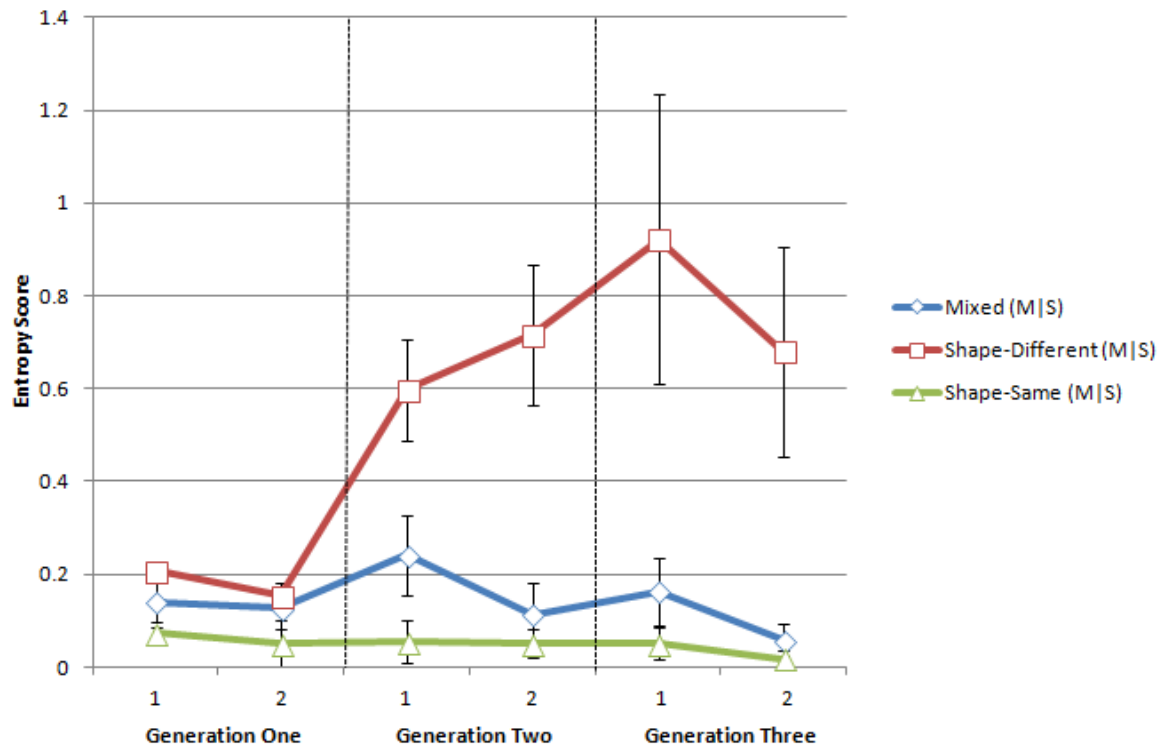
Figure Eleven

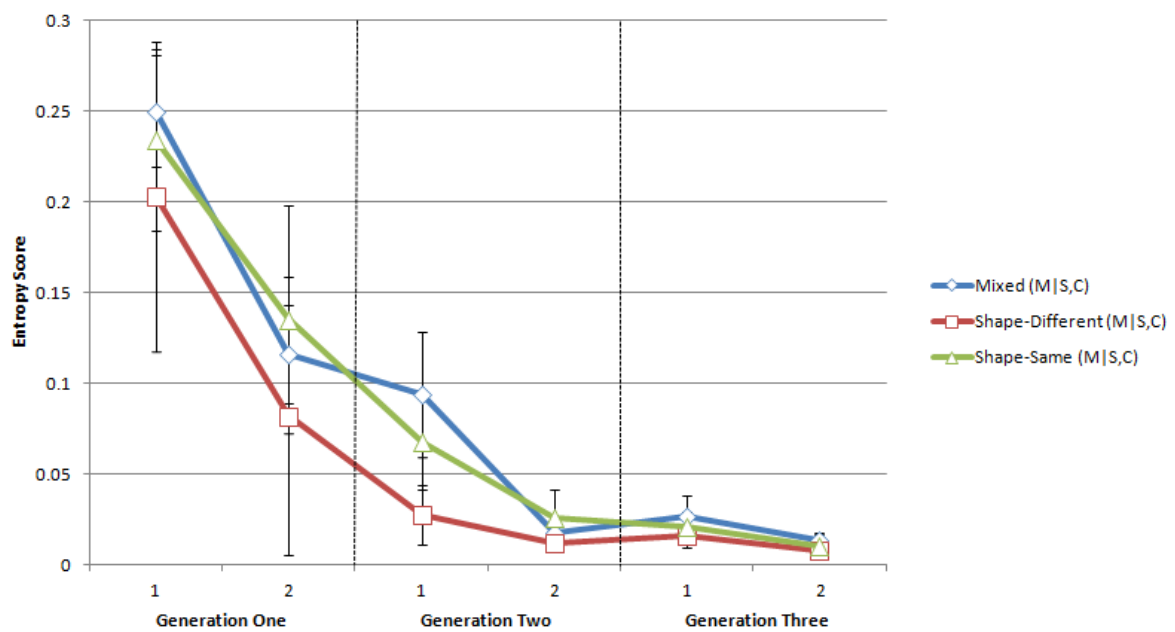Figure 11. Degree of meaning uncertainty, measured as H(M|S).

Figure Twelve

Figure 12. Meaning uncertainty of signals in context, measured as H(M|S,C).