

Achieving Compositional Language in a Population of Iterated Learners

Lewys Brace*, Seth Bullock, and Jason Noble

Institute for Complex Systems Simulation, University of Southampton, UK.

*L.G.Brace@soton.ac.uk

Abstract

Iterated learning takes place when the input into a particular individual's learning process is itself the output of another individual's learning process. This is an important feature to capture when investigating human language change, or the dynamics of culturally learned behaviours in general. Over the last fifteen years, the Iterated Learning Model (ILM) has been used to shed light on how the population-level characteristics of learned communication arise. However, until now each iteration of the model has tended to feature a single immature language user learning from their interactions with a single mature language user. Here, the ILM is extended to include a population of immature and mature language users. We demonstrate that the structure and make-up of this population influences the dynamics of language change that occur over generational time. In particular, we show that, by increasing the number of trainers from which an agent learns, the agent in question learns a fully compositional language at a much faster rate, and with less training data. It is also shown that, so long as the number of mature agents is large enough, this finding holds even if a learner's trainers include other agents that do not yet possess full linguistic competence.

Introduction

Human language is a learned system of symbolic representation that exhibits syntactic structure. Although the communication systems of other species appear to exhibit, at least to some degree, one or more of these features, the presence of all three in human language is arguably what makes it unique (Smith, 2002b).

Furthermore, human language has a number of notable design features, such as the way in which utterances are constructed from sub-parts, such as words and parts of words, which are reused and recombined in systematic ways. Thus, the meaning of an expression is related to the meanings of its constituent parts and the way in which they are combined. This trait enables language to be expressively open-ended, and is known as *compositionality* (Brighton and Kirby, 2001; Kirby, 2002b; Smith et al., 2003).

Kirby (2007) observes that compositionality endows human language with an obvious adaptive advantage in terms of its ability to communicate novel meanings; i.e., those that

have never before been expressed. Given the utility associated with the ability to construct a wide range of messages from just a few learned basic units (Kirby, 2013), it is remarkable that we do not see compositionality being used as part of a learned mapping between meanings and signals in the communication systems of other species¹.

The view that language is culturally-transmitted, and that this may have a crucial role in shaping the way in which it is formed (Smith, 2002a; Brighton et al., 2005; Christiansen and Chater, 2008) has led to a body of work arguing that compositional syntax may have arisen, not as a consequence of its utility to us, but because it better ensures the continued existence of the language itself (Kirby, 2007). This work sees the self-preservational development of language occurring as a result of a cultural-evolutionary process termed *iterated learning* (Brighton and Kirby, 2001; Kirby and Hurford, 2002; Smith et al., 2003; Kirby et al., 2008, 2014); the process whereby an individual learns their cultural behaviour from other individuals, who have themselves acquired their cultural behaviour in the same way. In other words, the input into an individual's learning process is, itself, the output of prior learning in other individuals.

Models of iterated learning and human language, then, involve an agent being presented with a set of meanings that it wishes to convey, choosing signals for each of these meanings, and then transmitting these meaning-signal pairs, or utterances, to another agent who then learns from them. This process is repeated generation after generation, and can be seen to represent how language competence and understanding can develop through observational learning (Brighton, 2002).

The distinction made within iterated learning research between the observable speech acts that fuel language learning on the one hand, and the individual's internal learned representation of a language on the other, is reminiscent of the concepts of *I-language* and *E-language* that were originally put forward by Chomsky (1986):

I-language: This is the pattern of neurons that implements

¹Although, as Kirby (2012) observes, bee dances do display limited compositionality.

an individual language user’s grammar within their mind.

E-language: This is the set of utterances that make up the spoken language.

Deacon (1997) argues that in order for language patterns to continue from one generation to the next, there is a requirement for a mapping from I-language to E-language and back again; he termed this the linguistic *bottleneck*.

The central contribution of the Iterated Learning Model (ILM) is first to have successfully idealised this process in a simplified setting that is amenable to study, and then to have demonstrated that the character of this bottleneck is crucially important to both whether or not language can be successfully passed from generation to generation and, in the situations where this transmission can be achieved successfully, show that it is also crucial to the character of the language that arises.

There have been numerous incarnations of the iterated learning model. Kirby, for example, has used versions of the ILM to look at the recursive properties of language (Kirby, 2002a) and compositionality (Brighton and Kirby, 2001). Hurford (2000) explores generalised phrase structure, while Brighton (2002) uses an ILM to explore the concept of the poverty of the stimulus (the fact that the data available to a language learner is sparse, yet the knowledge of language that they achieve is complex) and its relationship with a genetically coded innate language acquisition device.

Support for the ILM and the role of learner bias in language change has come about in recent years from both iterated learning experiments involving human participants, which have supported much of the work previously done with computational simulation (Kalish et al., 2007; Kirby et al., 2008), and from other methods of research such as the statistical analysis work of Lupyan and Dale (2010), who found that languages that are spoken by larger groups of individuals, such as modern English, tend to have simpler inflectional morphology² than those spoken by smaller groups.

It has even been suggested that the rarity of language in nature could, in part, be due to the rarity of iterated learning in the natural world (Kirby et al., 2014)³.

However, language learning in humans takes place within a complex social setting. Rather than each immature language user being assigned a single mature language user as a tutor, language users are exposed to linguistic input from a range of language users, some more mature than others.

This paper explores the changes to the behaviour exhibited by the ILM that result from situating language learning within a population of mature and immature learners. In the next section, we introduce an existing variant of the ILM,

²The process whereby adding a morpheme to a word either creates a different form of the word (i.e. car → cars) or a new word with a different meaning (i.e. car → caring).

³It is noteworthy, that certain species of songbird appear to learn their songs through a process akin to iterated learning.

and replicate its basic findings. We then describe an extended model featuring a population of learners and present results from this model. Finally we discuss the findings and conclude the paper.

The Iterated Learning Model

There have been several published variants of the Iterated Learning Model (ILM). Here we will extend one that was originally discussed by Kirby and Hurford (2002). It has four components:

1. A finite meaning space, \mathcal{M}
2. A finite signal space, \mathcal{S}
3. One speaker
4. One learner

Here, a language is defined as a mapping between a finite space of signals and a finite space of meanings. Each meaning and each signal are represented as an 8-bit binary string:

$$\mathcal{M} = \{m_1, m_2, \dots, m_{256}\}$$

$$\mathcal{S} = \{s_1, s_2, \dots, s_{256}\}$$

Each agent’s personal mapping from signals to meanings is implemented in the form of a three-layer feed-forward artificial neural network with eight nodes in each layer (see figure 1). Each of the eight nodes in the input layer is influenced by one of eight bits in an uttered signal. The degree of activation of each node in the input and hidden layers influences every node in the immediately downstream layer via a weighted connection. Each node’s activation is determined by the weighted input it receives from upstream nodes, squashed by a standard logistic activation function:

$$y_i = \frac{1.0}{1.0 + e^{-x_i}} + \theta_i + I_i$$

$$x_i = \sum_j \omega_{ji} y_j$$

Where y_i is the activation level of neuron, i , and x_i is incoming stimulation received by i , calculated as the weighted sum of upstream activation values. Each neuron also receives a constant bias input, $\theta_i = 1.0$, and may receive an external input $I_i \in \{0, 1\}$ if it is part of the input layer.

The activation values of the output layer are then translated into an 8-bit binary meaning by thresholding each node’s activation around the value 0.5. This string represents an agent’s best guess as to the meaning of the utterance that was input into the network. During learning, an agent updates the weights of its network using back propagation with a learning rate of 0.1 and no momentum term (Rumelhart et al., 1986).

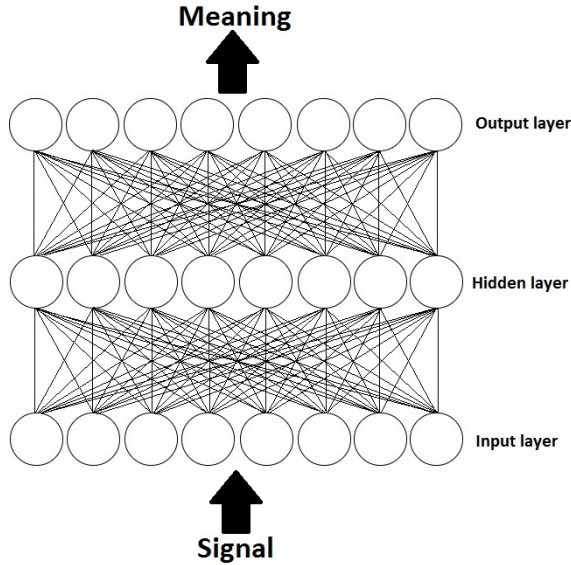


Figure 1: The agent’s neural network architecture.

Initially two agents are created, a mature language user (sometimes referred to as the “speaker”) and an immature language user (sometimes referred to as the “learner”). At the outset of the simulation there is no established language in place so, following Kirby and Hurford (2002), the mature language user is assigned a language comprising of a random mapping from each meaning to a randomly chosen signal. The immature language user is assigned a random neural network, i.e, each network weight is drawn from a normal distribution with zero mean and standard deviation 0.1, and each node’s bias input is 1.0.

The mature language user, M , then trains the immature language user, I , for a number of training episodes, T . Each episode involves M being assigned a meaning to express and generating an associated utterance, and I using their neural network to infer a meaning associated with that utterance. Any difference between the true meaning that M attempted to express and the meaning that I infers results in back propagation making changes to I ’s neural network in an effort to minimise this comprehension error. Note that in order for this supervised learning to take place, ILM models assume that I is able to make use of knowledge of the true meaning that M intended to convey.

The full set of training episodes that an immature language user experiences often comprises multiple exposures to the same fixed set of unique meanings. An agent might experience E epochs of training with each epoch comprising the same set of B randomly chosen unique meanings experienced in an order that is randomised for each epoch, i.e., $T = E \times B$. The number of different meanings communicated to a language learner, B , is referred to as the language learning “bottleneck”.

After all training episodes are complete, the mature language user is discarded, the immature language user is promoted to become the new mature language user, and a new randomly configured immature user is created to be trained. This process repeats for some fixed number of generations. Note that at the start of every generation the immature language user is assigned an entirely random neural network; there is no inheritance of language other than through experience of language learning episodes. Note also that the population structure is 1 + 1. At any moment in time one mature speaker is training one immature learner.

Since ILM agents have a neural network that maps *unidirectionally* from signals to meanings, they require an additional mechanism in order to generate signals for particular meanings. To this end, Kirby and Hurford (2002) adopt the *obverter* learning procedure that was originally formulated by Oliphant and Batali (1997). Here, each speaker assumes that the hearer’s internal mapping between signals and meanings is similar to its own and, consequently, when choosing which signal to make for a particular meaning, will choose the signal that, if presented as input to their own neural network, would most strongly cause them to infer this meaning, themselves. Oliphant and Batali (1997) prove that individuals using the obverter will tend to improve their communicative accuracy over time until an optimal communication system is achieved. Since the space of signals is finite and relatively small, this type of mechanism is feasible in the model.

In order to apply the obverter procedure within the ILM, Kirby and Hurford (2002) employ a confidence measure to determine which signal to produce for a given meaning. A speaker aiming to express a particular meaning, m , identifies their favoured signal, s^* , in the following manner:

For each signal, $s \in \mathcal{S}$, the speaker calculates an associated confidence value:

$$V_s = \prod_i (1 - |m[i] - o[i]|)$$

where $m[i]$ is the i^{th} bit of the target meaning and $o[i]$ is the i^{th} real valued output of the signaller’s neural network. The signaller then picks s^* as the signal with the largest confidence

ILM Results

We employ three metrics to evaluate language development, expressivity, stability and compositionality. A language’s *expressivity*, X , is the proportion of possible meanings that are generated by the full set of possible signals. A language with maximal expressivity is said to be complete. A language’s *stability*, S , is a relational property involving two agents and is measured as the proportion of the meaning space that can be recovered accurately when one agent signals to another. When a language is maximally stable, any meaning expressed by one agent can be inferred correctly by

the other.

The *compositionality*, C , of an agent’s language is the extent to which utterance parts convey distinct meanings. A language with zero compositionality is one in which every utterance is paired with a meaning in an uncorrelated fashion. Knowing part of the utterance provides no knowledge of part of the meaning. A fully compositional language is one in which every part of an utterance conveys perfectly an associated part of the meaning.

We evaluate the degree of compositionality in an agent’s language by first employing the obverter procedure to generate a signal for each of the meanings in the meaning space. We then calculate the values of each of the 8×8 correlations, C_{ij} , between the 256 values at the i^{th} bit of the set of signals and the 256 values at the j^{th} bit of the set of meanings. For each row, i , of this matrix we then calculate $C_{i*} = \max_j C_{ij}$, the maximum correlation between the values at index i of the signal set and the values at each of the indices of the meaning set. Finally, compositionality, C is calculated as the average of these eight maximal correlation values, $C = \frac{1}{8} \sum_i C_{i*}$. For a random language mapping meanings to signals, $C = 0.5$. Where a complete language is fully compositional, $C = 1$, each bit in an utterance conveys the value of one bit in the associated meaning.

The model displays three different types of behaviour, depending upon the size of the bottleneck. If the bottleneck is too small, then the agents do not learn; this results in a language that is both inexpressive and unstable. If, however, the bottleneck is too big, then an expressive and stable system is eventually reached; although, only after a prolonged period of time. Agents quickly achieve a language that is expressive and stable (see figure 2) and fully compositional (see figure 3) with a bottleneck of size 50.

Population-based Iterated Learning

The authors of the ILM themselves point out that complex population dynamics were traded off for computational power in the original model. Population structure was not taken into account, and every agent only ever learns from one other agent⁴. Given that the iterated learning model aims to shed light on the relationship between the properties of individuals and the population-level behaviour that they exhibit, and that much of the work done in this area thus far has been concerned primarily with vertical cultural transmission, it is of significant interest to explore the behaviour of this ILM within a population of agents.

Here we introduce a model in which, at each iteration, a population of N language users comprises N_M mature individuals and N_I immature individuals, where $N_M + N_I = N$. During each iteration of the model, every immature language user is assigned a number of trainers from whom they

⁴This is more than likely due to the computational capacities of the hardware at the time when the original model was developed.

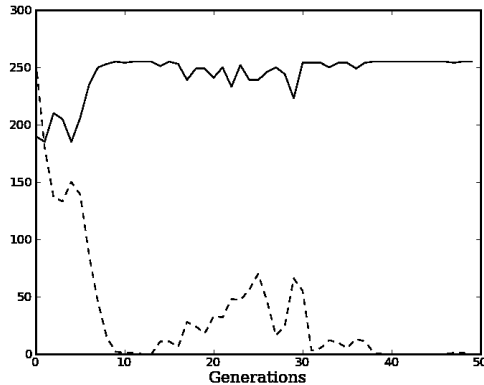


Figure 2: Replication of ILM behaviour. The solid line represents language expressivity, X , the proportion of the meaning space that is covered by the learner’s language. The dotted line represents the language instability, $256 - S$, the difference between the language mappings of the mature and immature language users. Here, $N_M=1$, $N_I=1$, $B=50$, $E=100$, $M_T=1$, $I_T=0$.

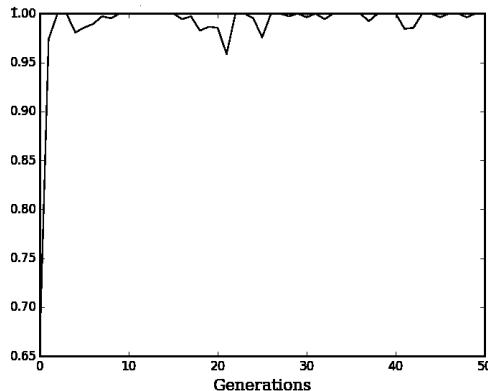


Figure 3: Language compositionality, C , over time for the ILM replication, where $N_M=1$, $N_I=1$, $B=50$, $E=100$, $M_T=1$, $I_T=0$.

infer the structure of their language through a series of training episodes. This set of trainers may involve both a number, M_T , of randomly chosen mature trainers, and also, possibly, a number, I_T , of randomly chosen immature trainers (see figure 4). The presence of immature trainers represents scenarios in which language learners are not kept isolated from one another, but may influence each others’ language learning. An immature individual’s total number of training episodes, T , is evenly split between their trainers with each trainer being involved in $\frac{B}{M_T + I_T}$ episodes per training epoch⁵. As in the original ILM, it remains the case that

⁵Fractional numbers of training episodes are avoided by round-

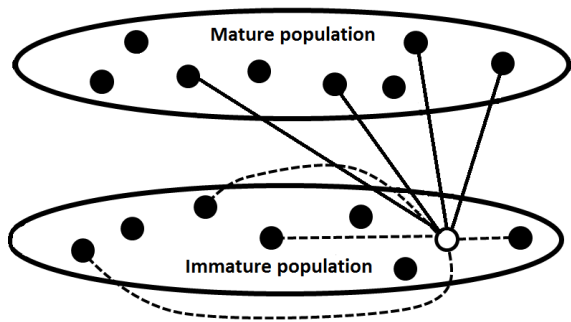


Figure 4: Diagrammatic representation of an ILM population divided into mature (upper set) and immature (lower set) agents, with $N = N_M + N_I = 16$ agents per generation. Lines represent one immature agent’s trainers: four mature trainers ($M_T = 4$, solid lines) and four immature trainers ($I_T = 4$, dashed lines).

the total number of training episodes, T is the product of the bottleneck size, B , and the number of training epochs, E . Hence, $T = B \times E$. The training episodes involving a specific trainer will involve the same set of randomly selected meanings in each training epoch. The set of B training episodes that comprise a single epoch are encountered in random order.

Iterated Learning Population Model Results

Figure 5 depicts a cross section of possible combinations of M_T and I_T , and how expressivity, X , and stability, S , develops in the population model.

In comparing figure 5A with figure 2, it is clear that a training input from multiple mature agents has a significant impact upon the number of generations required for a fully expressive and stable communication system to arise. Unsurprisingly, given the nature of iterated learning, figure 2B shows how the system fails to improve above the scores obtained by random chance when $M_T=0$. Figures 2C and 2D depict how the system is able to produce a largely expressive and stable system when both M_T and I_T are set equal, at 5 and 10, respectively.

To further explore the impact of multiple mature trainers on model behaviour, a series of tests were conducted with the aim of exploring the linguistic bottleneck. In figure 6, we see the result that different bottleneck sizes have upon compositionality in a population where $I_T=0$ and $E=50$; meaning that agents get half of the training sessions that they did in the original model, which should make learning far more difficult. The left graph showing $M_T = 1$ and the right showing $M_T = 10$. In both graphs, it can be seen that, when the bottleneck is set too low, the system does not learn. When agents learn from only one mature trainer, a bottleneck of at least 80 meanings is required

ing up.

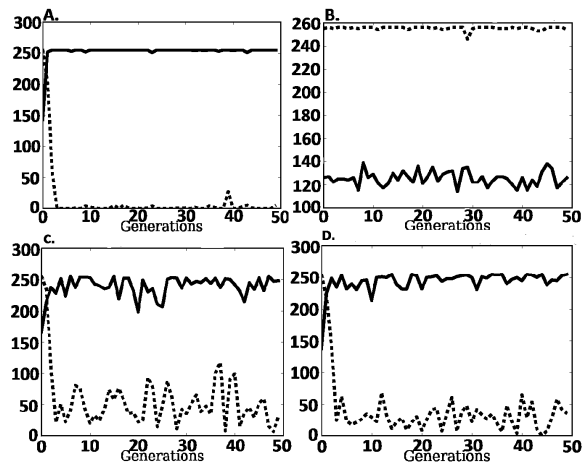


Figure 5: System behaviour for a single run of the ILPM simulation for various combinations of M_T and I_T . As above, the solid line depicts expressivity, X , and the dotted line represents instability, $256 - S$. Parameter settings are as follows: A. $M_T = 10$, $I_T = 0$; B. $M_T = 0$, $I_T = 10$; C. $M_T = 5$, $I_T = 5$; D. $M_T = 10$, $I_T = 10$; where $N_M=15$, $N_I=15$, $B=50$, and $E=100$ for all. Both the expressivity score and stability score are the average of the immature population after language learning has been completed.

before fully compositional language can survive. However, with ten mature trainers, a high level of compositionality can arise and survive with a much smaller bottleneck of around 50. Moreover, when compositional language arises, it does so far faster when multiple trainers are present.

Figure 7 depicts analogous results for scenarios in which immature language users are allowed to influence each others’ learning ($I_T = 5$). When immature trainers outnumber mature trainers (figure 7 left), language learning is compromised, with compositionality varying erratically over successive generations. Despite this, it is notable that bottleneck size does influence language with larger bottlenecks allowing languages to achieve somewhat higher compositionality. When immature trainers are outnumbered by mature trainers (figure 7 right), language learning is successful for scenarios with larger bottleneck sizes, although compositionality does vary more from generation to generation by comparison with an equivalent scenario without immature trainers (compare figure 6 right).

Further evidence of M_T impacting the system behaviour can be seen in figure 8, which plots the average level of compositionality that the system exhibits per generation for various combinations of M_T and I_T . In line with the above results, it can be seen that compositional language tends to arise to the extent that the number of mature trainers is greater than the number of immature trainers, and that a greater number of mature trainers enables the system to develop and maintain a higher level of language composition-

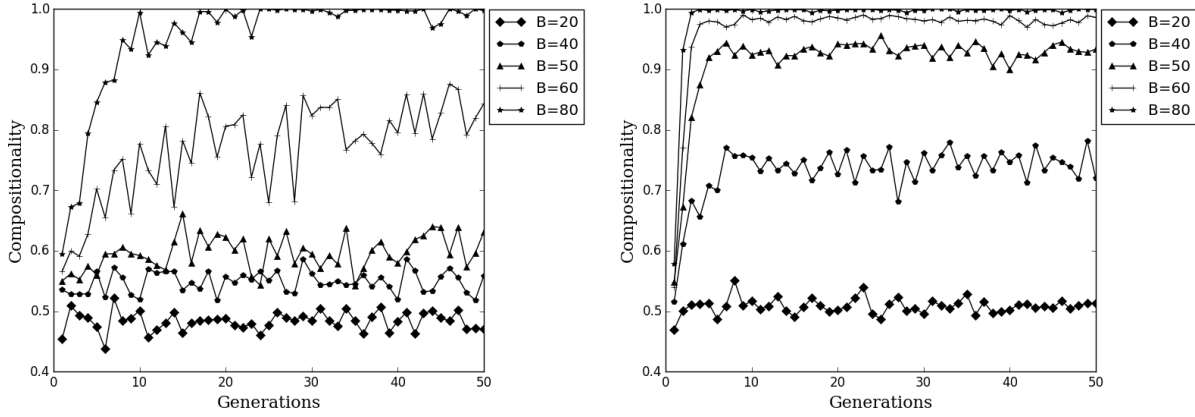


Figure 6: Graph depicting the impact of various value of B . *Left*: $M_T = 1$; *Right*: $M_T = 10$. ($N_M=15$, $N_I=15$, $I_T = 0$ and $E=50$ in both cases). The compositionality score is the average of the immature population after language learning has been completed.

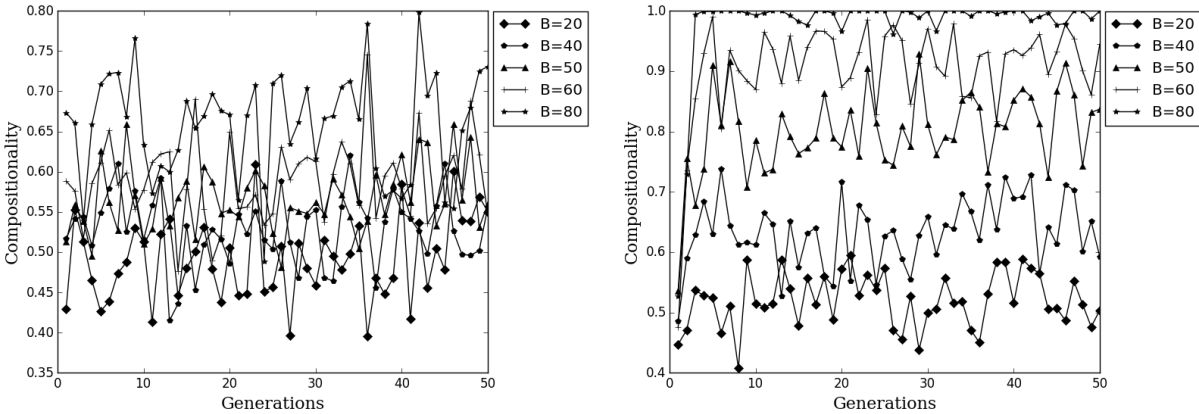


Figure 7: Graph depicting the impact of various values of B . *Left*: $M_T = 1$; *Right*: $M_T = 10$. ($N_M=15$, $N_I=15$, $I_T = 5$ and $E=50$ in both cases). The compositionality score is the average of the immature population after language learning has been completed.

ality.

Why might dividing the same number of learning episodes between a greater number of mature trainers lead to improved language learning in an immature language user? Several possibilities present themselves: multiple trainers could allow effective languages to spread through the population more quickly since one trainer can influence several learners, or, equivalently, expose learners to a sample of multiple languages, some of which may be more easy to learn and use. However, manipulating the population structure in ways that would be expected to influence this effect made no difference to performance.

Alternatively, might multiple trainers provide learners with increased diversity of language experience at the outset of the simulation, when naive neural networks tend to

map many meanings onto the same signal. Figure 9 lends some support for this hypothesis, showing that the number of unique signals experienced by a language learner at generation 2 of a run is increased when the learner is exposed to multiple trainers and that this increase in diversity is directly proportional to the increase in compositionality of the language exhibited a few generations later. However, it should be noted that although there was a strong relationship between signal diversity and compositionality across scenarios that differed in terms of the number of trainers, when the number of trainers was held constant there was not always a strong relationship of this kind, suggesting that signal diversity may not be the whole story.

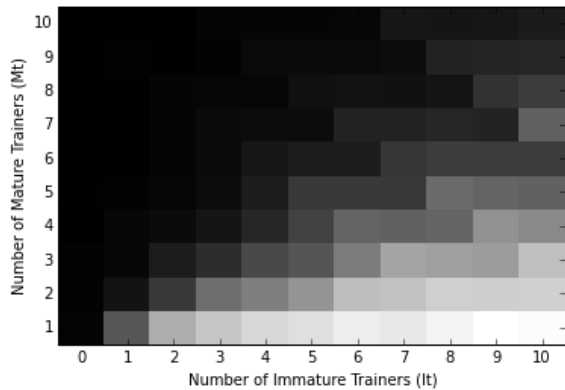


Figure 8: Heatmap of the average amount of compositionality over 50 generations, where $N_M=15$, $N_I=15$, $B=50$, and $E=100$, throughout. The compositionality score is the average of the immature population after language learning has been completed.

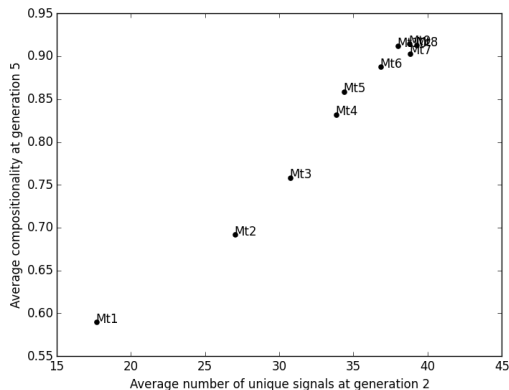


Figure 9: The relationship between the average number of unique signals that an immature learner experiences at generation 2 and the average compositionality of the language learned at generation 5 for runs with different numbers of mature trainers (M_T). Each data point represents an average over 10 runs where $N_M = 15$, $N_I = 15$, $B = 50$, $E = 50$.

Discussion and Conclusions

We have demonstrated that Kirby and Hurford’s (2002) iterated learning model variant can operate successfully within a population of agents. Given an appropriately sized language-learning bottleneck, when each member of a population of immature language users learn their language from enough mature language users, the population is readily able to converge on a complete, compositional language. Moreover, increasing the number of mature trainers tends to allow compositional language to pass through a smaller learning bottleneck and to establish itself in a smaller number of generations.

Work within the iterated learning paradigm typically holds that the way in which a language changes over time can be seen as a compromise the influence of learner biases and the influence of constraints acting upon language during transmission (Kirby, 2002a; Brighton et al., 2005; Smith, 2009). For example, the transmission bottleneck favours languages that can be inferred by language learners from a limited number of utterances (Kirby, 2002a; Brighton, 2002; Smith, 2009). Thus, the compositionality of a language represents an adaptation in response to selection pressures imposed by the environment in which it must survive. It is important to understand the dynamics of iterated learning within linguistic populations since population structure may be an additional source of constraints on language transmission and may therefore influence the form that languages tend to take over time.

The role of such constraints has been modelled previously in an iterated learning context. Griffiths (2007), for instance, explored iterated learning dynamics within a model where learning algorithms were based on the principles of Bayesian inference. By extending his framework to a population of such Bayesian agents where each learner learns from a single member of the previous generation, he showed that iterated learning in this population of Bayesian agents produced language outcomes that could be understood as solely the result of the agent’s individual learning biases. Therefore negating the role of other constraints, such as the transmission bottleneck.

However, Smith (2009) argues that Griffiths’ (2007) findings imply that it is possible to understand the prior biases of learners by looking at the typological distributions of languages. Smith (2009) also presents a model of Bayesian agents and demonstrates that Griffiths’ results are based upon the idealisation that a learner learns from a single teacher, and once multiple teachers are included, the mapping from the learner biases to typology breaks down. Based upon this result, Smith (2009) concludes that inferring learning bias from typology could yield unsafe results. Furthermore, Griffiths’ (2007) model is limited by the fact that the agents use very specific statistical learning algorithms, and are therefore not applicable to cases where the subjects of study use more general-purpose learning algorithms, which are more akin to the general purpose cognitive architecture that is likely to underpin human language (Hurford, 2014).

In a later work, Burkett and Griffiths (2010) explored the problems raised by Smith (2009) by developing a model where Bayesian agents were allowed to learn multiple languages. In doing so, they demonstrated that, so long as an agents hypothesis space explicitly takes into account the possibility of receiving input from multiple speakers with potentially different languages, then Bayesian learning does tend to reflect the learners inductive biases in the same manner as the single teacher model presented in Griffiths (2007). However, this model still makes the simplifying assumption

that agents only receive input from vertical transmission; this is clearly not the case for real-life language learners, who are likely to learn from their immature peers as well as from their mature role-models.

The model presented in this paper differs in this respect in that it explores the impact of immature language users upon the learning process, and the emergence of compositionality in particular. Furthermore, unlike Burkett and Griffiths (2010), we have explored iterated learning dynamics within a population of agents who are attempting to learn a single language. We have shown here that the introduction of horizontal language transmission amongst immature language learners does not tend to prevent languages from arising if each language learner is exposed to enough mature trainers.

Acknowledgements

The authors would like to thank Simon Kirby for his useful and friendly input. This work was supported by an EPSRC Doctoral Training Centre grant (EP/G03690x/1).

References

- Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54.
- Brighton, H. and Kirby, S. (2001). The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In Kelemen, J. and Sosík, P., editors, *ECAL 2001 - Full TITLE HERE*, pages 592–601. Springer-Verlag.
- Brighton, H., Kirby, S., and Smith, K. (2005). Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In *Language Origins: Perspectives on Evolution*. Oxford University Press, Oxford.
- Burkett, D. and Griffiths, T. L. (2010). Iterated learning of multiple language from multiple teachers. In *The Evolution of Language: Proceedings of the Eighth International Conference*, pages 58–65. World Scientific, Singapore.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Praeger Publishers, Westport, USA.
- Christiansen, M. H. and Chater, N. (2008). Language as shaped by the brain. *Behavioural and Brain Sciences*, 31:489–509.
- Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. Norton, New York, USA.
- Griffiths, T. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, 31:441–480.
- Hurford, J. R. (2000). Social transmission favours linguistic generalisation. In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge University Press., Cambridge.
- Hurford, J. R. (2014). What is wrong, and what is right, about current theories of language, in the light of evolution? *Humana Mente Journal of Philosophical Studies*, 27:123–133.
- Kalish, M., Griffiths, T. L., and Lewandowsky, S. (2007). Iterated learning : Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review*, 14(2):288–294.
- Kirby, S. (2002a). Learning, bottlenecks and the evolution of recursive syntax. In Briscoe, T., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, pages 173–204. Cambridge University Press., Cambridge.
- Kirby, S. (2002b). Natural language from artificial life. *Artificial Life*, 8(2):185–215.
- Kirby, S. (2007). The evolution of meaning-space structure through iterated learning. In Lyon, C., Nehaniv, Chrystopher, L., and Cangelosi, A., editors, *Emergence of Communication and Language*, pages 253–267, London. Springer-Verlag.
- Kirby, S. (2012). Language is an adaptive system: The role of cultural evolution in the origins of language. In *The Oxford Handbook of Language Evolution*. Oxford University Press, Oxford.
- Kirby, S. (2013). Language, culture, and computation: An adaptive systems approach to biolinguistics. In *The Cambridge Handbook of Biolinguistics*. Cambridge University Press, Cambridge.
- 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 of the United States of America*, 105(31):10681–6.
- Kirby, S., Griffiths, T., and Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28:108–114.
- Kirby, S. and Hurford, J. (2002). The emergence of linguistic structure an overview of the iterated learning model. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*. Springer, London.
- Lupyan, G. and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5(1):e8559.
- Oliphant, M. and Batali, J. (1997). Learning and the emergence of coordinated communication. *The Newsletter of the Center for Research in Language*, 11(1):1–46.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- Smith, K. (2002a). The cultural evolution of communication in a population. *Connection Science*, 14(1):65–84.
- Smith, K. (2002b). Natural selection and cultural selection in the evolution of communication. *Adaptive Behavior*, 10(1):25–45.
- Smith, K. (2009). Iterated learning in populations of Bayesian agents. In Taatgen, N. A. and van Rijn, H., editors, *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 697–702. Cognitive Science Society, Austin, TX.
- Smith, K., Kirby, S., and Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, 9(4):371–86.