

Research Article

Virve-Anneli Vihman*, Diane Nelson, Simon Kirby

Animacy Distinctions Arise from Iterated Learning

<https://doi.org/10.1515/opli-2018-0027>

Received December 2, 2017; accepted July 27, 2018

Abstract: Linguistic animacy reflects a particular construal of biological distinctions encountered in the world, passed through cultural and cognitive filters. This study explores the process by which our construal of animacy becomes encoded in the grammars of human languages. We ran an iterated learning experiment investigating the effect of animacy on language transmission. Participants engaged in a simple artificial language learning task in which they were asked to learn which affix was assigned to each noun in the language. Though initially random, the language each participant produced at test became the language that the subsequent participant in a chain was trained on. Results of the experiment were analysed in terms of learnability, measured through the accuracy of responses, and structure, using an entropy measure. We found that the learnability of languages increased over generations, as expected, but entropy did not decrease. Languages did not become formally simpler over time. Instead, structure emerged through a reorganisation of noun classes around animacy-based categories. The use of semantic animacy distinctions allowed languages to retain morphological complexity while becoming more learnable. Our study shows that grammatical reflexes of animacy distinctions can arise out of learning alone, and that structuring grammar based on animacy can make languages more learnable.

Keywords: animacy, iterated learning, learning bias, cognitive bias, artificial language learning, language evolution

1 Introduction

Most, if not all, human languages are sensitive to the animacy of noun referents on some level. Animacy distinctions frequently show up as a feature in morphology and syntax, and animacy often conditions choices speakers make about word order or choice of synonymous constructions, even where it is not systematically encoded. Yet linguistic animacy is not a simple reflection of biological distinctions encountered in the world, but rather it is mediated by a particular construal of those distinctions, passed through cultural and cognitive filters of what is most important to humans, such as conspecificity and empathy, sometimes also employing other relevant properties such as texture and edibility. We are interested in the process by which our construal of animacy becomes encoded in the grammars of human languages.

Article note: This paper belongs to the special issue on Effects of Animacy in Grammar and Cognition, edited by Diane Nelson and Virve Vihman.

***Corresponding author: Virve-Anneli Vihman**, Institute of Estonian and General Linguistics, University of Tartu, Jakobi 2, 51014 Tartu, Estonia, E-mail: virve.vihman@ut.ee

Diane Nelson, Linguistics and Phonetics, School of Languages, Cultures and Societies, University of Leeds, Leeds LS2 9JT, UK

Simon Kirby, Centre for Language Evolution, Dugald Stewart Building, 3 Charles St., Edinburgh EH8 9AD, University of Edinburgh

Animacy is conditioned by categories such as agency and intentionality, which are key to structuring events and transitive sentences. However, the cross-linguistic effects of animacy on grammar are not always a reflection of its importance for event structure, but rather suggest that animacy has a deeper resonance in human cognition and culture. If animacy informs our perception and construal of the world in such fundamental ways, then its role in grammars may also be more fundamental: animacy distinctions may emerge and retain their role in grammar not only for communicative purposes, but also for enabling language transmission. If speakers have a learning bias which benefits from categories based on animacy distinctions, then that bias may provide a mechanism for how animacy becomes encoded in grammars, and simultaneously explain the ubiquity of animacy effects in language. In this study, we asked whether animacy has an effect on the learnability of language, using an iterated learning paradigm.

The structure of the paper is as follows. In the next section, we introduce linguistic animacy and briefly discuss ways in which it diverges from binary classes and biological animacy. Following this, in Section 3, we introduce the methodology used in our study. In Section 4 we present our results, analysing the learnability and structure of the languages which arose during intergenerational transmission in the study, and in Section 5, we discuss the implications and conclusions.

2 Linguistic animacy

As a conditioning factor in morphosyntax, animacy can be reduced to a binary distinction between animate (alive, sentient) and inanimate (non-living, non-sentient) referents, as witnessed in differential object marking (DOM) in Spanish (1a-b, see, e.g. von Heusinger & Kaiser 2011: 600-601).

1. a. Carlos besó a su amiga. (animate)
 Carlos kiss.PAST at his friend
 ‘Carlos kissed his friend.’
- b. Carlos besó [Ø] su retrato. (inanimate)
 Carlos kiss.PAST his picture
 ‘Carlos kissed his picture.’

Cross-linguistically, however, we find that animacy is often treated as a much more nuanced, gradient category. The sorts of distinctions made vary widely, and have been subsumed under various scales, schematised in the General Animacy Scale shown in (2) to represent the hierarchical nature of linguistic animacy (see Lockwood & Macaulay 2012, Yamamoto 1999, 2006).

- (2) General Animacy Scale
 Human > Animal > Inanimate > Abstract

Linguistic animacy hierarchies typically encode humans at the top end of the scale, followed by animals, inanimate objects and abstract entities. This scale generalises over the many ways languages carve up these categories, as the granularity of animacy distinctions varies across languages. Moreover, it has been noted that linguistic animacy is not always aligned in a straightforward mapping onto biological animacy (see Cherry 1992, de Swart & de Hoop 2018, Bayanati & Toivonen, this issue).

The human category, for instance, may be further subdivided, for example according to discourse participants, with first and second person higher than third person. Nouns referring to adults may be treated differently from nouns for children (e.g. German: morphologically diminutive, hence neuter *Mädchen* ‘girl’ and *Fräulein* ‘young/unmarried lady’ vs feminine *Frau* ‘woman’).

Animals, too, are often divided into further, culturally relevant subcategories such as large vs small, or tame vs wild, as discussed by Yamamoto (1999). The Manam language of New Guinea, for instance, makes use of dual and paucal forms (and concomitant verb agreement patterns) only for humans and ‘higher animals’, a category containing tame, anthropomorphised animals such as dogs, pigs, birds, and in the

contemporary language, also goats and horses (Croft 1990: 113, Lichtenberk 1983: 110, cited in Yamamoto 1999).

Likewise, the animacy continuum posited for Navajo affects verb form and word order, making the distinctions shown in (3) (Young & Morgan 1987: 65-66):

(3) Human > Infant/Big Animal > Medium-sized Animal > Small Animal > Natural Force > Abstraction

Note that in (3), human infants are distinguished from other humans, and grouped instead with large animals: this semantic classification is clearly not based on visual similarities such as size or shape, but reflects a relative scale based on attributed animacy as relevant to human life, even cutting across species distinctions. Note also that natural forces form a distinct category of their own in Navajo.

Although we see more differentiation on the high end of the hierarchy, inanimate objects and abstract entities may also vary greatly in cultural significance and other properties relevant to animacy, such as self-propelled motion (consider cars or windmills) and attributed sentience (machines, computers, telephones). Plants are usually grouped with inanimates, but they grow, consume nutrients, and breathe.

Animacy is a perceptually salient and evolutionarily useful category for humans to employ (Dahl 2008), and its cognitive construal is complex, hierarchical and gradient. At the same time, it has been proposed that the grammatical encoding of linguistic animacy tends to invoke discrete, binary features or oppositions which bisect the animacy hierarchy at certain points (de Swart & de Hoop 2018). For example, DOM systems conditioned by animacy involve a binary marked vs unmarked case option, as in Spanish (1), while animacy-based word order systems like Navajo and Sesotho operate over a restricted set of grammatical functions, typically subject vs object. The encoding of animacy in grammars, then, involves an asymmetry in complexity: the conceptually complex and gradient nature of animacy does not directly map onto formal complexity (for characterisations and discussion of the principles of linguistic complexity, see Audring 2017, Di Garbo 2016, Kusters 2003, Miestamo 2008). But the ubiquity of animacy distinctions in grammars shows that there are functional pressures for animacy to be introduced into formal systems and retained over time. We are interested in exploring this tension between semantic complexity on one hand and formal complexity on the other.

Our proposal in this paper is to look at animacy as a cognitive bias shared by humans. We ask whether including animacy as a linguistic feature affects the learnability of a language, potentially allowing languages that are more complex to survive the process of cultural transmission – a process that normally leads to simpler systems (Culbertson & Kirby 2016). If animacy does affect learnability in this way, then languages may respond adaptively by encoding animacy distinctions. Although both communicative function and learnability are key drivers of language change (Kirby *et al.* 2015), here we propose that the pressure from learning alone is sufficient. We conducted an experiment to investigate the effects of animacy on the learnability of a toy language transmitted across generations.

3 Method

The experiment used a standard iterated learning paradigm (Kirby, Cornish & Smith 2008; Kirby, Griffiths & Smith 2014) to explore the effect of cognitive biases on the cultural transmission of a miniature language with noun class marking¹. Participants engaged in a simple artificial language learning task where they attempted to learn which of three nominal markers is assigned to each of 24 different nouns. Each participant was assigned to one of 20 transmission chains such that the language produced by a participant at test in a chain became the language that the subsequent participant in the same chain was trained on (Figure 1). In

¹ Languages which incorporate animacy via noun class or gender systems tend to assign class membership via verbal agreement morphology and other means rather than marking on the noun itself. In fact, in the context of markedness and indexing, Malkuchov (2008:210) notes that “it is redundant to mark animacy per se (as it is clear from the lexical content of the nominal in question).” Our toy grammar therefore represents a typologically unusual language; however, marking animacy on the noun lends itself to a straightforward experimental design.

this way, only the initial languages for each of the 20 chains was created by us, the experimenters. We refer to the sequence of steps in these transmission chains as “generations” in the experiment.

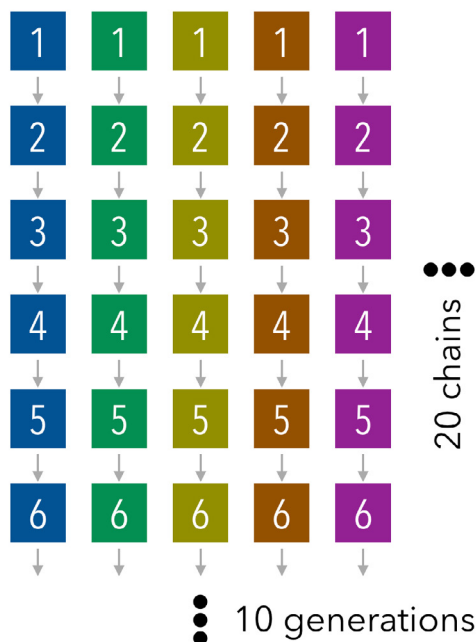


Figure 1: The iterated learning design. Each participant is assigned to one of 20 chains. They learn the language produced by the previous participant in their chain, with the exception of the first participant in a chain, who learns a randomly constructed language. In this way, each participant belongs to a generation, and we observe the cultural evolution of the language down 10 generations of iterated learning.

3.1 Participants

200 English-speaking participants were recruited using the online crowdsourcing platform, Crowdfunder. They were each paid \$1.50 for participation, and the experiment took approximately 10 minutes to complete. Ethical approval was granted for this experiment according to the procedures of the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh. All participants provided informed consent and were offered debriefing information.

3.2 Stimuli

Participants learned the use of three nominal affixes in a miniature artificial language by observing sentences in this language of the form: an initial intransitive verb, followed by a noun with an affix attached. This verb-subject order was used to make the language syntactically unlike English. Each sentence was presented in the artificial language with a gloss in English for the verb and subject underneath. No gloss was provided for the noun affixes. The verb was glossed as “appears” in every case (we used this verb since it is compatible with nouns of all animacy classes). The artificial language sentence was highlighted in red, and the noun affix was additionally marked in bold and attached to the noun with a hyphen. In addition, a black-and-white line drawing depicting the subject noun was shown above the sentence. See Figure 2 for an example stimulus presentation.



Figure 2: Example training (left) and testing (right) screens. In training, participants are shown an image of the noun, the artificial language, and a gloss for the verb and noun. In testing, their task is only to recall which affix goes with the noun being presented.

All syllables used in every language had a Consonant-Vowel structure. The monosyllables were used for the three noun-class markers, and the words for the verb and nouns for each language were chosen randomly from the set of bisyllabic forms. See Table 1 for a complete list of the possible monosyllabic and bisyllabic forms used in the experiment.

Table 1: The set of monosyllables and bisyllables used in the experiment

Monosyllables	Bisyllables
bo, ta, ki	homa, hopa, hopo, howu, leho, lele, leli, lema, lemi, lene, lepa, lepo, lewu, liho, lile, limi, lipa, lipo, liwu, maho, mapa, mapo, mawu, miho, mima, mimi, mipa, mipo, miwu, neho, nele, neli, nema, nemi, nene, nepa, nepo, newu, paho, pama, pami, papo, pawu, poho, poma, pomi, pone, popa, popo, powu, wuho, wule, wuli, wuma, wumi, wune, wupa, wupo, wuwu

The 24 nouns used in the experiment were chosen to represent a range of different possible sub-classes of relevance to animacy distinctions found in the world's languages. Specifically they were: *adult, teenager, child, baby, teddy, robot, horse, sheep, rabbit, bear, fish, clam, amoeba, flower, cactus, tree, fire, lightning, wind, rain, school, car, rock, love*. See Figure 3 for the complete set of images used. Visual cues represented a potential confound, so the size, salience and directionality of the images were made as similar as possible. These cross-cut a range of possible animacy distinctions. For example, there are 4 humans, 9 vertebrates (animals including humans), 2 living invertebrates (*amoeba* and *clam*), 3 plants, 4 natural forces, and 2 motile inanimate objects (*car* and *robot*). *Teddy* is inanimate but animal in form, while *robot* is motile, humanoid in form, and potentially construed as intelligent or sentient.



Figure 3: The full set of images used in the experiment.

3.3 Design

Each of the 20 chains in the experiment was “seeded” with a different, random, initial language which the first participant in each chain tried to learn. Each random language was constructed in such a way as to ensure that each of the three noun affixes was used for exactly 8 nouns, but the selection of which noun was presented with which marker was completely random. The verb form and each noun form was randomly selected from the list of bisyllabic words, and the form of each noun affix was similarly selected at random from the monosyllabic forms.

Subsequent languages in the chain were created from the previous participant’s output in the test phase of the experiment with the additional step that the actual *form* of the verb, nouns, and noun-class markers was reassigned randomly at each generation of the experiment. This randomisation process effectively ensures that there can be no lasting effect of the phonological form of the words on the languages that emerge. For example, any phonological (in)compatibility between noun and affix at one generation will not be present in the next. Rather, what is transmitted across generations is the grouping of nouns into noun classes labelled with distinct affixes.

Similarly to other iterated learning experiments, we look at two aspects of the experimental results. First, we examine how the languages’ *learnability* changes as a function of generation. Typically, in iterated learning experiments, we expect learnability to increase over time. In other words, participants in later generations should have lower rates of error in the task: their choice of noun affixes for each noun should mirror the noun affix choices made by the previous participant at test.

Secondly, we look at *structure* in the languages. The initial input languages are unstructured by design, having been created by random assignment of affixes to nouns. We predict that any increase in learnability at later generations will have arisen because the languages will have become more structured. We expect after an initial period of evolution away from randomness, the chains will be exploring a space of structured languages that reflect the cognitive biases of our participants. Specifically, we predict that the distribution of noun classes over meanings in these late-generation languages will reflect animacy distinctions. An alternative outcome would be the simplification of languages by making the affixes more predictable overall, irrespective of the animacy of the noun. This could be done by skewing the distribution of markers such that one is much more common than the others, or in a more extreme case, the removal of one or even two affixes, to create a simplified grammatical structure. In approaches which seek to measure grammatical complexity (Audring 2017, Di Garbo 2016), a system with two affixes to signal a particular function is less complex than a system with three affixes. This reduction of complexity through an increase in predictability and loss of informative distinctions is a common finding in iterated learning experiments, which normally is only avoided by having some kind of communicative pressure in play (Kirby et al. 2015). Note that we are not including any kind of communicative pressure in this experiment by design, but instead predict that animacy distinctions alone will prevent the reduction in complexity of the affix distributions that would otherwise be expected from iterated learning.

3.4 Procedure

The experiment was conducted via the participants’ own web browsers. Participants were assigned to one of 20 chains at random until a chain had reached 10 generations. They were either exposed to one of the initial, random languages if they were the first participant in a chain, or to the languages generated from the previous participant in the chain for subsequent generations. The experiment proceeded in two main phases: the training phase and the testing phase.

In the training phase, participants were exposed to 18 randomly chosen items out of the full 24 items in the language. There were three blocks of training, each of which involved training on all 18 items. Each block consisted of 6 “mini” blocks of training. In a mini block, three items were presented one after the other, followed by a mini test of recall for one of those three items. The training schedule was constructed in such a way that every one of the 18 items was presented three times, and was the subject of a mini test once.

In the training presentation, the image for the current item was shown on the screen for 1 second, after which the artificial language sentence and its English gloss was shown below the image for a further 3 seconds. No input was required from the participant, and the next training trial was presented automatically. For the mini test, the image was shown with the artificial language sentence and gloss, but with the noun affix left missing. Three buttons were shown with each of the three possible markers on them and the participant was invited to click on the correct button. The order of the buttons was randomised across trials.

In the testing phase, participants were presented with all 24 items once (i.e. both the 18 that they saw in training, plus the additional 6 unseen items). The testing proceeded exactly like the mini test trials in the training phase.

4 Results

We will analyse the results of the experiment in terms of *learnability* and *structure* in turn. Recall that we expect learnability to increase over generations, even in the absence of communication, and structure to emerge in the languages later in the experiment, such that the distribution of noun markers will reveal animacy-based cognitive biases of our participants, even though some of these distinctions are not present in English.

4.1 Learnability

We measure the learnability of a language by counting the number of times (out of a maximum of 24) that the participant trained on that language used the correct noun affix. Figure 4 shows how the learnability of languages in the experiment increases over generations. We used a linear mixed effects model in R (R Core Team 2012) and *lme4* (Bates, Maechler & Booker 2013) to analyse the relationship between learnability and generation. We entered a fixed effect of generation and a random intercept for chain, and by-chain random slopes for the effect of generations. Using a likelihood ratio test to compare the full model with a reduced model without the fixed effect revealed that generation affected the learnability of languages ($\chi^2(1)=12.56$, $p<.001$).

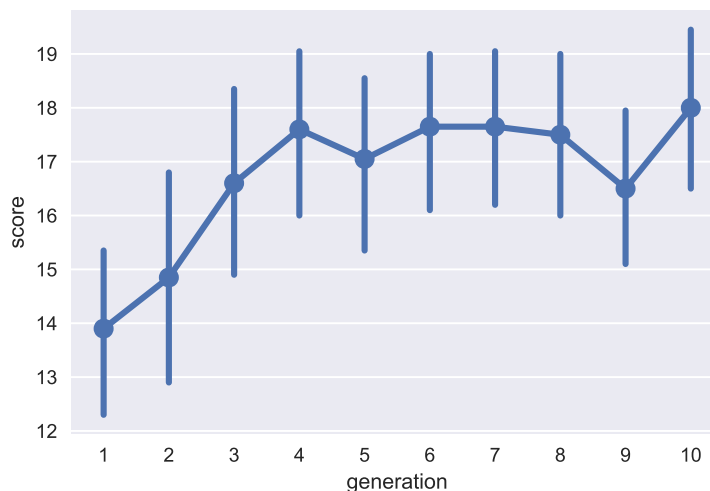


Figure 4: The number of correct responses made by participants in testing (out of a maximum of 24) over generations. Bootstrapped 95% confidence intervals are shown. Accuracy increases over generations.

4.2 Structure

One way in which languages may become more learnable is by simply reducing the structural variability of noun affixes to make the marking more regular and predictable. In the limit, our languages could lose noun markers altogether, and in this way make the learning task easier (as predicted by characterisations of ‘relative complexity’; see e.g. Kusters 2003). This is a familiar pattern in a range of iterated learning experiments (Kirby, Cornish & Smith 2008) and can be measured by calculating the entropy of the language (Smith & Wonnacott 2010; Ferdinand, Kirby & Smith 2017). Entropy reflects not only how many markers there are, but also their distribution. If one marker is more common than the others, then the entropy is lower than if all markers are equally common (which give the maximum entropy possible). Entropy therefore gives a measure of how predictable or regular markers are in general.² Simpler, more regular languages are expected to be more learnable, and therefore typically arise as a result of iterated learning. To test whether this would explain our results, rather than the emergence of animacy-based distinctions, we looked at entropy of the languages as a function of generation (Figure 5).

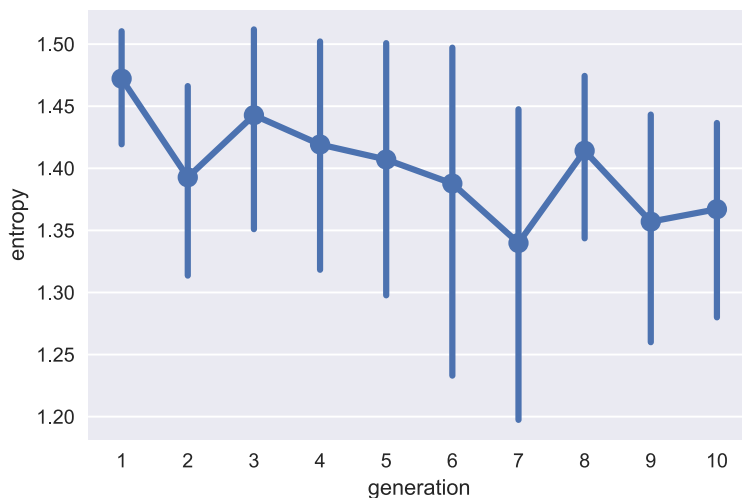


Figure 5: The entropy of the output language produced by participants in testing over generations in bits. Bootstrapped 95% confidence intervals are shown. The input language at the start of every chain has the maximum possible entropy of 1.58 bits (with each of three noun class markers used for exactly 8 nouns). The maximum entropy for a language with 2 classes would be 1 bit. Although there appears visually to be a slight downward trend, this is not significant: counter to expectation, the languages do not become structurally simpler over generations.

We analysed entropy using a linear mixed effects model with a fixed effect of generation and a random intercept for chain, and by-chain random slopes for the effect of generations. A likelihood ratio test comparing the full model with a reduced model without the fixed effect revealed that generation did *not* affect the entropy of languages ($\chi^2(1)=2.71$, $p=0.10$). In other words, unlike in other iterated learning experiments (e.g., Reali & Griffiths 2009), languages at the end of the experiment were not significantly simpler (i.e. more predictable in their use of markers) than those at the start. This does not appear to be an artefact of our experimental design. Although all three possible markers are present as options for participants, nevertheless occasionally the number of markers reduced (with a consequent loss in entropy). In other words, this was not ruled out by the experimental design. Nevertheless, this only happened 4 times

² To put it another way (and equivalently), entropy is a measure of how much expected information a marker contains. The more markers there are, and the more uniform their distribution, the higher the average amount of information about possible meanings you could recover from knowing the marker.

(out of 200 participants) in the course of the study. More importantly, as we showed above, there was no significant downward trend in entropy. We will return to this in the discussion.

Since the languages are not becoming more learnable by virtue of any decrease in entropy, there must be some other way in which they are becoming structured. We can get a sense of what this is by looking at an example language from the data. Table 2 shows the final language of one of the chains in the experiment. We can see here a system of noun affixes that seems to be based at least in part on animacy, with *-ta* being used exclusively for human nouns, all animals marked with *-bo* and all natural forces marked with *-ki*. Inanimates and borderline nouns are found with both *-bo* and *-ki*.

Table 2: An example language produced at the end of a chain in the experiment, with nouns grouped by markers used. This language shows a clear effect of animacy, with *-ta* used exclusively for humans, and *-ki* being used for natural forces and three of the other non-animals.

Class marker	Nouns
<i>-ta</i>	adult, teenager, child, baby
<i>-bo</i>	robot, horse, sheep, rabbit, bear, fish, clam, amoeba, flower, cactus, school, car, love
<i>-ki</i>	teddy, tree, fire, lightning, wind, rain, rock

For a more global picture of the results, we looked at the distribution of noun affixes in the experiment after the effect of the initial random languages had been washed out by several generations of cultural transmission. Kalish, Griffiths & Lewandowsky (2007) demonstrate how, under certain assumptions about the learning process, transmission chain experiments such as this one can be used to reveal participants' biases. This is because, after enough time has passed to wash out the effect of the initial languages in the chain, iterated learning is predicted to converge to the prior bias of the learners. Here we are treating the first half of the chains as a "burn in" period to allow for the effect of the random initial languages to decay and the prior biases to be revealed. Of course, it is possible that 5 generations is not long enough, and the evolving languages have not converged to a distribution shaped solely by the cognitive bias of our participants. However, this is a conservative assumption, since any effect of the initial random languages will merely add noise to our results rather than structure. For these purposes, we consider only the 100 languages produced in the second half of each chain in the experiment. We used two methods for visualising the distribution of noun classes in the experiment: *multi-dimensional scaling*, and *hierarchical cluster analysis*.

The first stage in this analysis involved creating a distance metric between pairs of nouns based on all the languages in our sample. For each pair of nouns, we calculated the proportion of languages in our sample that used a different noun-class marker for those two nouns. A distance of zero would indicate that every one of our 100 languages uses the same marker for the two nouns in question; conversely, a distance of 1 would indicate that every language assigns the two nouns to different classes.

This set of distances implies a space of noun meanings with very high dimensionality. We reduce these down to two dimensions using the isoMDS function in the MASS package in R. Figure 6 shows the resulting position of the nouns in the 2d space found by the MDS algorithm. Several clusters can be seen clearly in this diagram that relate to animacy: humans, land animals, plants, and natural forces, for example. These clusters themselves appear to be organised on a larger dimension of something akin to "aliveness" from roughly right to left in the plot, with the natural forces being least "alive" and humans most alive. In addition, although the dimensions in an MDS plot are not meaningful, nevertheless some orthogonal sub-hierarchies are apparent in the plot. Within each cluster there appears to be a vertical dimension which we might describe as "potency", cross-cutting the "aliveness" distinction, yielding sub-hierarchies such as: adult > teenager > child > baby, and wind > fire > lightning > rain.

This dimension may have emerged out of finer-grained cognitive biases related to animacy. It is compatible, for instance, with the results of a study by Cherry (1992), in which adult participants were asked to use intuitive judgements to rank pairs of nouns according to the categories of embodiment (does it have a body, related to physical size), purposiveness (which conflates sentient "sense of purpose" with utility), and activity (equivalent to motility). Independent from rankings of aliveness, subhierarchies emerged

within biologically equivalent pairings; for example *redwood* was ranked higher than *daisy* for embodiment and purposiveness; *wind* outranked *thunder* for purposiveness and activity; and *tiger* outranked *hippo* for all three. These look intriguingly similar to the relationships which emerged in the subdomains in our MDS analysis, with tree above flower, wind above lightning and horse above rabbit and sheep. No firm conclusions can be drawn from this impressionistic view, but the compatibility of results from studies of very different kinds is suggestive and merits further study.

An alternative method for clustering the data is to use hierarchical cluster analysis on the distance matrix between nouns. Figure 7 shows the result of clustering using the Ward method (from the linkage function in the Python SciPy cluster library). The major split that this method brings out of the data is the one between higher animates (including humans) vs all other nouns, but also visible are the clusters of humans and natural forces, seen in the MDS analysis as well. *Teddy* appears to be treated as a higher animate in this analysis, clustering alongside *bear*.

Finally, we subjected the final languages in the experiment to a test of whether they reflected animacy distinctions to a degree greater than expected by chance. Mutual information provides a measure of how much we can learn about one variable by observing another. We can use this to test whether the distribution of noun affixes predicts the animacy of the nouns they are attached to (and vice versa). We measured the mutual information between the noun classes and four binary categories (human, higher animates, plant, natural force) and summed these for the final languages in the experiment, giving us a grand total over all 20 languages of 9.87 bits. A permutation test showed that this was higher than would be expected by chance, $p < 0.0001$. We compared the mutual information sum from the experimental data with 10,000 simulated experiments in which the noun-class proportions of each language were the same as the ones found in the real experiment but the assignment to nouns was randomly permuted. None of the permutations yielded a summed mutual information score across 20 simulated languages that was higher than 9.87 bits.

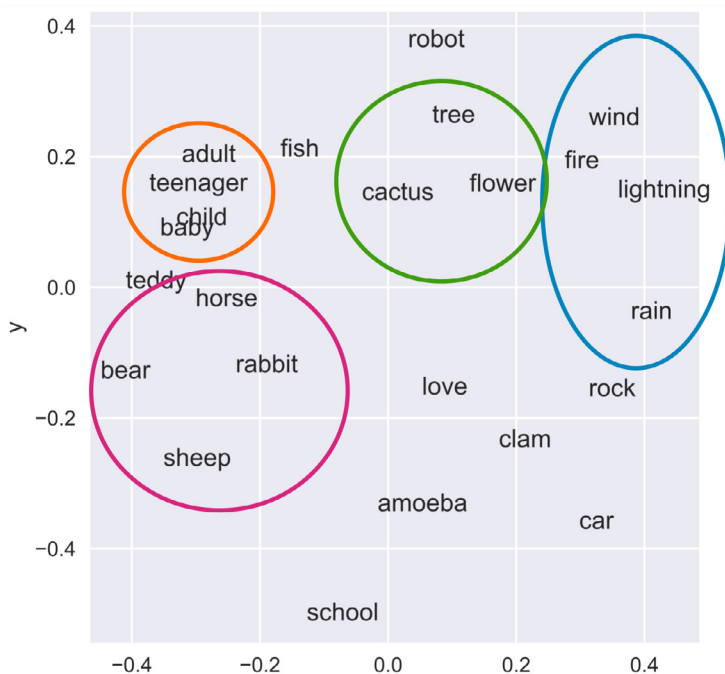


Figure 6: A multidimensional scaling (MDS) plot showing the distances between the 24 nouns, averaged over the 100 languages in the second half of the experiment. Two nouns are close in this plot if they tend to take the same noun-class marker within a language, and far apart if they tend to take different noun-class markers. Four animacy-related clusters (humans, plants, land animals, natural forces) have been circled for illustrative purposes.

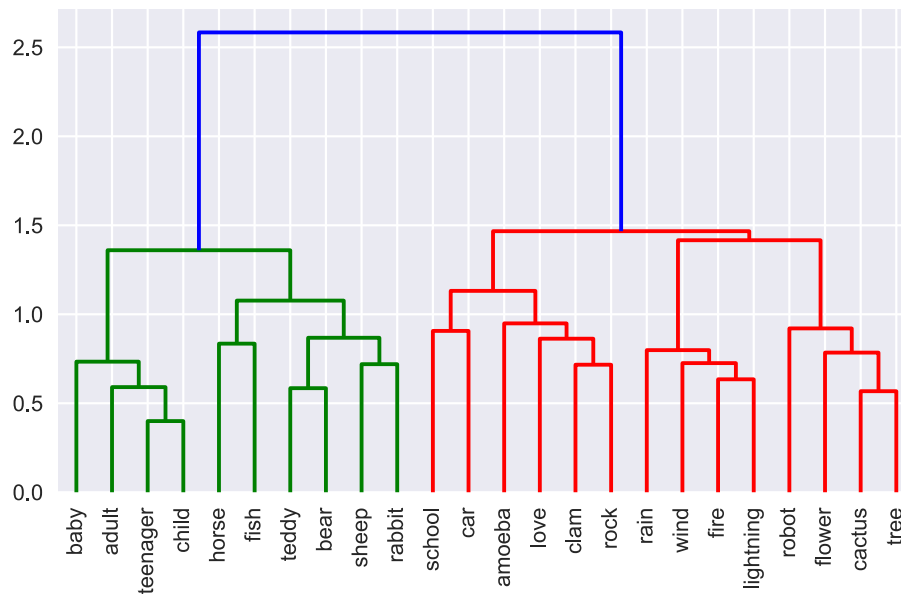


Figure 7: Hierarchical cluster analysis of the similarity between nouns in the 100 languages in the second half of the experiment. As in the MDS plot in Figure 6, nouns are clustered together to the extent that they tend to bear the same noun affix. The major split shown here is between vertebrates (humans and higher animals) and others (invertebrates, inanimates and abstract nouns).

5 Discussion

We ran an iterated learning experiment in order to investigate the effect of animacy on language transmission. Participants engaged in a simple artificial language learning task in which they attempted to learn which of three noun-class markers was assigned to each noun in the language. Each participant was trained on 18 of 24 nouns and tested on all 24 nouns. Participants were each assigned to a transmission chain, such that the language the participant produced at test in a chain became the language that the subsequent participant in the same chain was trained on.

We analysed the results of the experiment in terms of learnability and structure. Learnability was expected to increase over generations, and semantically based structure was expected to emerge in languages at later points in the chains in the experiment. Learnability of a language was measured through the accuracy of the responses for that language, and we analysed the relationship between learnability and generation using a linear mixed effects model. We found that the learnability of languages in the experiment increased over generations, as expected.

The initial languages were created by random assignment of affixes to nouns, and hence were unstructured. After an initial period of evolution away from randomness, we expected the chains to explore a space of structured languages that reflect the cognitive biases of our participants. To investigate whether languages became more learnable by reducing structural variability and increasing regularity, we measured the entropy of the languages as a function of generation. Although entropy often decreases in iterated learning experiments, we found no effect of generation: languages were not becoming simpler or “more regular” over time. Languages could, in theory, reduce the numbers of markers being used: in a handful of cases, a language emerged with only two markers. However, looking across all chains and generations, the overall pattern was not one of accumulating loss of variability. Instead, we found that structure emerged not through simplification via reduction, but through a semantic reorganisation of noun classes around animacy-based categories as the languages were transmitted through iteration over the course of ten generations.

Our experiment deliberately starts with a language with complex, unstructured, unconditioned variation and we see that instead of reducing formal variation, iterated learning reorganises the system to take advantage of animacy distinctions. How do we interpret the lack of decrease in entropy and the emergence of animacy distinctions in our paradigm? Languages generate complexity in various ways due to a variety of historical processes (Kusters 2003, Nichols 1992). Iterated learning tends to reduce this complexity – it removes variation because this makes language more learnable. However, there are ways in which complexity can be retained: specifically, if it is conditioned by something that learners can easily pick up. This allows iterated learning to retain richness in a system where it might otherwise be removed by the tendency toward simplification.

Languages that are more predictable are likely to be easier to learn, which is why a common finding in iterated learning experiments is for variation of various kinds to be lost. For example, in the first experiment in Kirby, Cornish & Smith (2008) the number of distinct words rapidly reduces over 10 generations. Reali & Griffiths (2009) and Smith & Wonnacott (2010) build on artificial language learning experiments by Hudson Kam & Newport (2005) and show how unpredictable variation is lost over generations, leading to regularisation in language. Ferdinand, Kirby & Smith (2017) look in detail at this regularisation process and show how it can be modulated by features of the task. Kirby et al. (2015) and Culbertson & Kirby (2016) suggest that the results of these iterated learning experiments make sense in the light of a highly general bias for compressible hypotheses in learning (Chater & Vitányi 2003). Essentially, formally simple languages are easier to learn, so the process of iterated learning inevitably leads to simplification. In the case of our experiment, the simplest language would be one in which only a single affix is used for all nouns, but instead we see the languages with the same complexity (in terms of the variability in noun affixation) at the end of the experiment as at the beginning. In previous experiments (e.g., Kirby, Tamariz, Cornish & Smith 2015) this retention of variability in the language comes from a pressure for communicative utility working against the bias for simplicity. However, in our experiment there is no such communicative pressure. Participants are not using the language to communicate, and in any case, the noun affix is strictly redundant given the presence of the noun. Nevertheless, the languages in the experiment do indeed become easier to learn over generations.

We suggest that in this case, animacy is providing an alternative semantic route to increased learnability that reduces the pressure on the language to jettison formal variation in marking. This is somewhat akin to the finding in Smith & Wonnacott's (2010) study in which variability in an artificial language is retained over generations but that variation shifts from being completely unpredictable to being completely conditioned by context. Similarly, Carstenson, Xu, Smith & Regier (2015) in an iterated learning study looking at the use of spatial language see the retention of multiple forms over many generations as the use of those forms comes to map onto spatial concepts in a structured way. This suggests that although a language which does not use noun markers at all will be easier to learn (and of course such languages exist), there will be less pressure on an evolving language to erode the use of such markers in the case where their distribution ends up being conditioned by a semantic feature, in this case animacy.

This also explains why relatively complex grammaticalised systems involving animacy as a conditioning feature are retained over time cross-linguistically (see Nichols 1992, Corbett 1991). For example, in a sample of 84 African languages exhibiting grammatical gender, Di Garbo (2016:67) found that languages with purely semantic gender systems (in contrast to phonologically- or morphologically-based formal systems or mixed systems) tended to invoke animacy as the relevant semantic dimension. Our results suggest that an elaborated set of animacy-related features in the grammatical structuring of a language does not significantly add to the complexity of that language for learners (in the sense of 'relative complexity'; see Kusters 2003), because conditioning variation on animacy distinctions reflects our learning bias.

Our results also support Comrie's (1989:186) cross-linguistic observation that "...animacy can be a relevant parameter in language change even where it is not particularly salient in the synchronic state of the language prior to the change." The same cognitive bias may underlie the role of animacy effects in diachronic patterns of restructuring noun class systems (also known as grammatical gender systems). The prototypical Bantu noun class system, for instance, is not primarily based on animacy, although typically humans are assigned to the same class. Nonetheless, studies have shown that in cases where a Bantu

noun class system undergoes restructuring, animacy is likely to play a role (Verkerk & Di Garbo 2017, Wald 1975), and animacy-based restructuring is also attested in other language families. Igartua & Santazilia (this issue) discuss evidence from multiple language families that shows how animacy can constrain the morphological complexity of inflectional paradigms in diachronic processes: morphological complexity is retained when animacy structures the semantic complexity underlying it. This is highly compatible with our findings from a toy language, and suggests that the learning bias we postulate, based on animacy, can be found to function in real examples of language transmission and change.

Animacy is a salient feature of the world around us, and a useful one for humans to recognise and attend to, in order to recognise mates and conspecifics, or potential threats. One might argue, however, that visual properties such as size and shape are equally (or more) salient. In fact, noun class systems may include both animacy and visual properties, as for instance the noun class system of Gújjolaay Eegimaa, an Atlantic language spoken in Senegal (related to the Bantu systems mentioned above), which has a noun class devoted to humans, as well as distinct classes based on size and shape (Sagna 2012). Future studies could investigate stimuli in which animacy scales are pitted against visual or phonological cues. Our study did not show that animacy is a better conditioning factor than any other, as our stimuli were chosen to reflect animacy categories rather than other possible classification systems. Rather, we showed that when a grammar uses animacy as an organising principle, that alone can make the language more learnable, and hence the process of iterated learning can lead to the reflection of animacy distinctions in language.

6 Conclusion

In our study, we showed that languages in an iterated learning paradigm became more learnable through cultural transmission without decreasing in entropy. We found that animacy distinctions roughly consistent with biological taxonomies can make languages more learnable when used as organising principles in grammar. Animacy was enlisted as an organising principle for the noun affixes in the languages our participants learned. The use of animacy distinctions made it possible for a language to retain complexity while also becoming more learnable: the language was not formally simplified or regularised, but it was restructured around semantic principles.

Our study also showed, importantly, that animacy distinctions can arise out of learning without communicative pressure. In other words, learning biases alone are enough to lead to the inclusion of animacy distinctions in language, without the need to express concepts in a comprehensible way to a listener. Our task tapped only learning and retention, rather than communication. Animacy-based systems can emerge based on learning biases alone.

Acknowledgements: We gratefully acknowledge funding from the European Union's Seventh Framework Programme (Marie Curie IEF grant no 623742 to first author), which helped fund part of this research. We also thank two anonymous reviewers for their helpful comments, and the audience at the 14th International Cognitive Linguistics Conference in Tartu for useful questions. Many thanks to Tartu-based artist Katrin Kelpmann, who drew the stimuli used in the experiment, and to the participants in the study.

References

- Audring, Jenny. 2017. Calibrating complexity: How complex is a gender system? *Language Sciences* 60. 53–68.
- Bayanati, Shiva, Ida Toivonen. This issue. Humans, animals, things and animacy features. *Open Linguistics*.
- Carstensen, Alexandra, Jing Xu, Cameron T. Smith, Terry Regier. 2015. Language evolution in the lab tends toward informative communication. In Noelle, David, Rick Dale, Anne Warlaumont, Jeff Yoshimi, Teenie Matlock, Carolyn Dacey Jennings, & Paul Maglio (eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 303–308. Austin, TX: Cognitive Science Society.
- Chater, Nick, Paul Vitányi. 2003. Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Science* 7(1). 19–22.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology* (2nd ed). Chicago: University of Chicago Press.
- Croft, William. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.

- Culbertson, Jennifer, Simon Kirby. 2016. Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology* 6. 1964. doi: 10.3389/fpsyg.2015.01964.
- Dahl, Östen. 2008. Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua* 118. 141–150.
- Di Garbo, Francesca. 2016. Exploring grammatical complexity crosslinguistically: The case of gender. *Linguistic Discovery* 14 (1). 46–85.
- Ferdinand, Vanessa, Simon Kirby, Kenny Smith. 2017. The cognitive roots of regularization in language. arXiv preprint arXiv:1703.03442.
- Hudson Kam, Carla, Elissa L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1:2. 151–195. doi: 10.1080/15475441.2005.9684215.
- Igartua, Iván, Ekaitz Santazilia. This issue. How animacy and natural gender constrain morphological complexity: Evidence from diachrony. *Open Linguistics*
- Kalish, Michael L., Thomas L. Griffiths, Stephan Lewandowsky. 2007. Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review* 14 (2). 288–294.
- Kirby, Simon, Hannah Cornish, Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105 (31).10681–86.
- Kirby, Simon, Tom L. Griffiths, Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* 28. 108–114.
- Kirby, Simon, Monica Tamariz, Hannah Cornish, Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141. 87-102. doi: 10.1016/j.cognition.2015.03.016.
- Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflections*. Dissertation. Utrecht: LOT, University of Leiden.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Lichtenberk, Frank. 1983. *A Grammar of Manam*. Honolulu: University of Hawaii Press.
- Lockwood, Hunter, Monica Macaulay. 2012. Prominence hierarchies. *Language and Linguistics Compass* 6/7. 431–446. doi: 10.1002/lnc3.345.
- Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In Miestamo, Matti, Kaius Sinnemäki, Fred Karlsson (eds.), *Language Complexity: Typology, contact, change*, 23–41. Amsterdam: John Benjamins.
- Realí, Florencia, Thomas L. Griffiths. 2009. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111. 317–328.
- Sagna, Serge. 2012. Physical properties and culture-specific factors as principles of semantic categorisation of the Gújjolaay Egimaa noun class system. *Cognitive Linguistics* 23 (1). 129–163. doi: 10.1515/cog-2012-0005.
- Smith, Kenny, Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116 (3). 444–449.
- de Swart, Peter, Helen de Hoop. 2018. Shifting animacy. *Theoretical Linguistics* 44 (1-2). 1–23. doi: 10.1515/tl-2018-0001.
- Verkerk, Annemarie, Francesca Di Garbo. 2017. Correlates of restructuring in Bantu gender systems. Paper presented at 50th Annual Meeting of the Societas Linguistica Europaea, Zürich, Sept. 2017.
- Von Heusinger, Klaus, Georg. A. Kaiser. 2011. Affectedness and Differential Object Marking in Spanish. *Morphology* 21:593–617. doi: 10.1007/s11525-010-9177-y.
- Wald, Benji V. 1975. Animate concord in northeast coastal Bantu: Its linguistic and social implications as a case of grammatical convergence. *Studies in African Linguistics* 6: 267–314.
- Yamamoto, Mutsumi. 1999. *Animacy and Reference: A cognitive approach to corpus linguistics*. Amsterdam: John Benjamins.
- Yamamoto, Mutsumi. 2006. *Agency and Impersonality: Their linguistic and cultural manifestations*. Amsterdam: John Benjamins.