Authors
Prasertsom, Ponrawee
Smith, Kenny
Culbertson, Jennifer

Domain-general categorisation principles explain the prevalence of animacy and absence of colour in noun classification systems

Ponrawee Prasertsom, Kenny Smith, Jennifer Culbertson
({ponrawee.prasertsom, kenny.smith, jennifer.culbertson}@ed.ac.uk)
Centre for Language Evolution, University of Edinburgh, UK

Abstract

Animacy is prevalent as a semantic basis for noun classification systems (i.e., grammatical gender, noun classes and classifiers), but colour is completely absent, despite its visual salience. The absence of colour in such systems is sometimes argued to suggest domain-specific constraints on what is grammatically encodable. Here, we investigate whether this tendency could instead be explained by the superior predictive power of animacy (i.e., the degree to which it predicts other features) compared to colour. In a series of experiments, we find that animacy-based noun classes are learned better than colour-based ones. However, when participants are encouraged, by manipulating predictive power, to sort images based on colour, they are subsequently worse at learning animacy-based noun classes. The results suggest the animacy bias in grammar may have its roots in domain-general categorisation principles. They further serve as evidence for the role of cognitive biases in constraining cross-linguistic variation.

Keywords: animacy; colour; gender; noun class; typology; artificial language learning; domain-specificity

Introduction

Despite their great diversity, languages exhibit recurring patterns, sometimes called *language universals* (Croft, 2002; Newmeyer, 2005). This study is concerned with a universal of noun classification systems (Aikhenvald, 2017). These are grammatical systems found in many languages, which divide nouns into a set of classes, in some cases indicated by morphological agreement on other words (e.g., grammatical gender and noun class), and in others only indicated by co-occurring markers or morphemes (e.g., classifier systems).

Noun classification is often based, at least partly, on the semantics of the noun. The two most common semantic features relevant in such systems are social gender or perceived sex (Kramer, 2020) and *animacy* (Corbett, 2013). For example, Swahili and other Bantu languages have animacy-based noun classes, reflected through prefixes on both the noun and agreeing words. In contrast with these two features, there are other semantic features which appear not to be used as a basis for classification. For example, no known language has colour-based noun classification. This may be surprising, given that colour categories are highly salient in other domains of cognition. In vision, Holmes and Regier (2017) found that English speakers exhibited lateralised categorical perception between *warm* and *cool* colours, being able to discriminate e.g. between yellow and green better than blue and green in the right visual field. This absence of colour and pervasiveness of animacy in noun classification, and other grammatical domains, has long been noted in theoretical linguistics research, and has led some to propose that only certain conceptual domains, of which animacy is one, are made available to language learners during acquisition (Adger, 2018; Cinque, 2013; D'Alessandro, 2021; Talmy, 1988).

A domain-specific constraint on grammar is not, however, the only possible explanation. The bias for animacy in noun classification could be driven by a domain-general principle that governs categorisation, both linguistic and non-linguistic. Aikhenvald (2000) suggested that common semantic bases for classification correspond to conceptual domains that yield high within-category similarities and inter-category differences, i.e., high *cue validity* (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Lee, 1988). These categorial (dis)similarities allow humans to predict other features from existing ones. In the categorisation literature, Bayesian approaches derive categorisation preferences from the predictive power of features (Anderson, 1991; Griffiths, Sanborn, Canini, Navarro, & Tenenbaum, 2011). Simplicity-based approaches likewise suggest that the most informatively compressed categories are highly predictive (Pothos & Chater, 2002; Pothos, Chater, & Hines, 2011). Under this view, animacy is more common as a basis for noun classification because it is highly predictive, while colour is not. Intuitively, it is possible predict from animacy a number of other features of an object: the ability to move, the organic nature, the possession of limbs and so on. On the other hand, knowing colour is much less helpful: knowing that an object is warm-coloured doesn't tell you much else. This latter view is consistent with much recent experimental research, according to which language universals arise due to domain-general cognitive biases active during learning and processing (Culbertson & Kirby, 2022; Culbertson & Newport, 2015; Kemp & Regier, 2012; Kirby, Tamariz, Cornish, & Smith, 2015; Maldonado, Zaslavsky, & Culbertson, 2023).

Here, we test the hypothesise that the bias for animacy over colour as a basis for noun classification is ultimately due to the difference in predictive power. We first test whether there is such a bias at all in an artificial noun class learning task, taking accuracy in learning as a proxy for the bias (Exp 1). We then test whether predictive power underlies the bias (Exp 2a-2b). We manipulated the predictive power of colour and animacy features and tested whether learning cor-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Figure 1: The set of visual stimuli used in Exp 1. Each image is either animate (*frog*, *lizard*) or inanimate (*present*, *bag*), and is cool- (*blue*, *green*) or warm-coloured (*red*, *yellow*).

relates with predictive power. To preview, the results of Exp 1 confirmed the animacy bias. In Exps 2a-2b, we did not find straightforward evidence for the predicted effects—likely due to the strength of the a priori bias for animacy. However, exploratory analyses showed that the change in predictive power modulated the animacy bias indirectly. Participants whose preference in a sorting task was shifted toward colour by its higher predictive power subsequently learned animacy-based noun classes worse than they would have.

Exp 1: Animacy- vs. colour-based noun classes

Materials

We taught participants an artificial language consisting of 16 nouns and two definite determiner variants (*da* and *te* 'the'). Each noun is a combination of an entity type (animate: frog, lizard; inanimate: bag, present) and a colour (warm: red, yellow; cool: blue, green). In order to simplify learning, each noun's syllables were designed to resemble their English counterparts (e.g., *frigru* for 'green frog'). The audio for each word was generated using MacOS TTS (`say`) with the voice Samantha. We used a set of 16 images to represent these meanings (Figure 1).

Methods

Each participant was assigned to one of the two conditions (animacy or colour). The experiment consists of two main phases: noun learning and noun class learning (via determiners), followed by a post-experiment questionnaire. In the noun learning phase, in each trial, participants were shown an image from the set accompanied by two buttons, each of which contained a noun in the language. Participants listened to the audio of the noun, and had to click on the button matching it. The trial was repeated if the response was incorrect. Each noun-image was presented twice, hence 32 trials in total. After initial training, they were further trained to criteria. The trials were exactly the same, except no audio was played and participants had to recall the noun from the image. They proceeded to the next phase after scoring at least 13 in 16 consecutive trials.

In the noun class learning phase, participants were told that the language had two words for 'the' and were instructed to find a pattern for when each is used. They then learned the determiners the same way they did the nouns, except each image now had a word with a blank space (e.g. ___ *frigru*), and the audio of the full phrase rather than the noun alone was played. Participant had to click on the determiner matching what they heard, and feedback was provided. The determiner varied based on either the noun's colour (e.g. *da* for *warm*, *te* for *cool*) or animacy (e.g., *da* for *animate*, *te* for *inanimate*), depending on the condition. At test, the trials were the same, except the audio played only contained the noun, and participants had to choose the correct determiner. No feedback was provided. Each unique training trial was presented 3 times (for a total of 48), and each test trial was presented twice (for a total of 32). Finally, participants were asked to give their strategy in choosing the determiner, and to provide information about their language background.

The determiner-category mappings, the noun-determiner relative order, and the order of trials in each phase were randomised per participant. This and subsequent experiments were all implemented with jsPsych (de Leeuw, Gilbert, & Luchterhandt, 2023).

Participants

To minimise L1 influence, we targeted native speakers of English, which does not have a noun classification system, and whose grammar makes relatively little reference to animacy (i.e., only in singular third person pronouns). We exclude participants who indicated non-negligible knowledge of languages with noun classification on the post-experiment questionnaire, and who scored lower than 90% on the noun train trials, the vocabulary testing trials, or the determiner training trials. All participants in this and subsequent experiments were recruited through Prolific (`https://prolific.co`) and compensated at the rate of 9 GBP per hour. We kept collecting data until we reached 80 participants after exclusion (40 per condition).

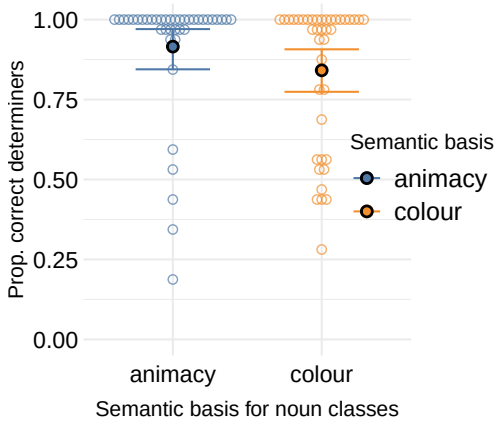

Figure 2: Proportion of correctly selected determiners (*y*-axis) for Exp 1 participants (unfilled dots) by semantic basis (*x*-axis). Error bars are 95% CI around the means (filled dots).

Results

Our main prediction for Exp 1 is that, if an animacy bias is at work during noun class learning, participants should score higher in the determiner learning test in the animacy condition. Consistent with this prediction, Figure 2 shows that participants in the animacy condition outperformed those in the colour condition (Mean prop. correct answers = 0.915 for animacy vs. 0.841 for colour). We fit a logistic regression model predicting per-trial correct responses from condition with by-participant and by-noun random intercepts, and a by-noun random slope on condition. A likelihood ratio test reveals a significant improvement in model fit when condition is included as a fixed effect ($\beta_{\text{semBasis}} = 1.859$; $p = 0.023$; deviation-coding). In summary, while participants in both conditions were generally very good the learning both systems, they are better at learning an animacy-based system.

Exp 2a: Manipulating predictive power

The results from Exp 1 have shown that there is a small but significant bias for animacy over colour in a simple artificial noun class learning task. Recall that above, we proposed (following e.g., Aikhenvald, 2017) that this bias is due to the fact that animacy features are more predictive of other features than colour. In Exp 2, we tested whether this mechanism can explain better learning of animacy-based noun classes in our task. To do this, we first manipulate the predictive power of animacy and colour in our stimuli. We predict that participants should learn animacy-based noun classes better when animacy is more predictive than colour; by contrast they should learn colour-based classes better when colour is more predictive than animacy. In other words, learning should be facilitated when the more predictive feature serves as the basis of classification—i.e., when predictiveness is *congruent* with semantic basis.

Materials

We used 16 meanings and 2 'the' variants (*da* and *te*) as in Exp 1. However, in order to full control meaning dimensions, here meanings corresponded to unfamiliar objects that differ along 5 dimensions (animacy, colour, shape, horn type, and appendage type). Noun labels were drawn from those used in Fedzechkina and Jaeger (2020); i.e., the forms do not hint at meaning, but still conform to English phonotactics. The audio of each word was generated as in Exp 1. The label-meaning mappings were randomised per participant.

To manipulate predictive power, we created two sets of image stimuli that differ only in whether colour or animacy is more predictive of the other features. Higher predictive power is given by lower conditional entropy. For example, if animacy is more predictive, then $H(F_i|F_A) < H(F_i|F_C)$, for every feature aside from animacy and colour F_i, where F_A and F_C are animacy and colour features. To ensure symmetry, the same feature matrix is used for both sets, the sole difference being which feature is associated with higher predictive power. The matrix also ensures that animacy/colour is the most predictive of the features aside from animacy and colour (i.e., shape, horn type, appendage type). Figure 3 shows the set of stimuli where animacy is more predictive.

Methods

We manipulated whether the most predictive feature in the stimuli (animacy or colour) was used as the basis of classification or not. Each participant was therefore assigned one of four conditions that differ in whether 1) animacy or colour is more predictive of other features and 2) animacy or colour is the basis for noun classification. In the two congruent conditions, either animacy or colour was both most predictive, and the basis of classification; in the two non-congruent conditions, the feature that was most predictive was *not* the one used for classification.

Based on piloting, we determined that using a image sorting task prior to learning was an effective method of familiarizing participants with the predictive structure of the stimuli. Therefore, the experiment consisted of two phases: familiarisation and noun class learning. The familiarisation phases started with an image sorting task. Participants were presented with all the stimulus images in two rows, in a random order, and were instructed to sort all of them into two boxes appearing below (Figure 4). The participants were free to sort them as desired, but must sort all of the images into the two boxes to proceed.

After this, participants were exposed to the same stimulus images once more, one at a time. In each trial, an image was shown. After 1.5 seconds, a button appeared allowing participants to proceed to the next trial. Every 4 trials, there was an attention check, where participants had to choose an image they had just seen from two choices.

After familiarisation, participants were trained and tested on the noun classes using the same procedure as Exp 1.

Participants

We excluded all participants who scored less than 90% on the attention check trials or the determiner train trials. We kept collecting data until we reached 240 participants after exclusion (60 per condition).

Results

To determine whether a bias for more predictive features could be observed in a non-linguistic (sorting) task, we first look at the image sorting responses in the familiarisation phase. Figure 6 shows the proportion of best-fit sorting strategies for participants separated by whether animacy or colour was the most predictive feature. The best fits were determined by the following method. We calculated the adjusted mutual information (Vinh, Epps, & Bailey, 2010) between each participant's sorting and reference sorting categories (i.e., based solely on animacy, colour, horn type, etc.). We took the reference sorting that yielded the highest information as the sorting strategy of a given participant.

As Figure 6 shows, while participants sorted by animacy the most regardless of predictive features, they were sig-

Figure 3: The set of images in Exp 2a-2b in the conditions where animacy is more predictive. Note, for example, that animates (stimuli with faces) are circular, have a crescent-like horn, and have squiggly appendages; most inanimates are inkblot-shaped, have a zigzag horn and have straight appendages.

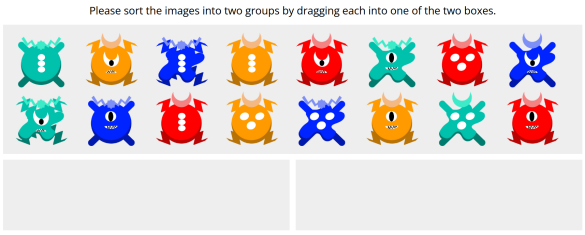

Figure 4: The image sorting phase in Exp 2a-2b. Colour features are more predictive in the stimuli depicted.

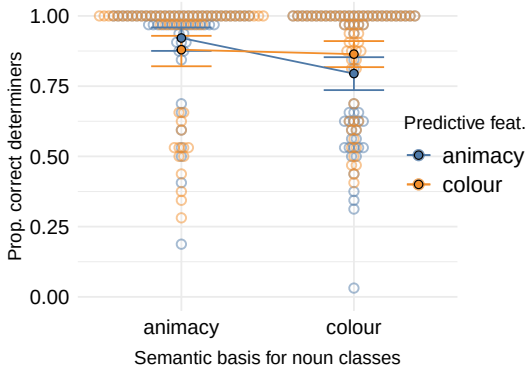

Figure 5: Proportion of correctly selected determiners (y-axis) for Exp 2a participants (unfilled dots) by semantic basis for noun classes (x-axis) and most predictive feature (dot colour). Error bars are 95% CI around the means (filled dots).

nificantly more likely to sort by colour when colour was more predictive than animacy (12.5% vs. 26.667%; $\chi^2(1) = 6.773, p = 0.009$). They were also significantly more likely to sort by animacy when animacy was more predictive than colour (50.833% vs. 36.667%; $\chi^2(1) = 4.334$, $p = 0.037$).

Moving on to the noun class learning results, Figure 5 shows the proportion of correct determiner responses for participants based on the semantic basis of the system (animacy or colour), and which feature is most predictive (animacy or colour). Recall that when animacy is the most predictive feature (blue dots), we predicted animacy-based classes to be learned better; when colour is the most predictive feature (orange dots), we predicted colour-based classes to be learned better. To test our prediction, we first fit a logistic model predicting per-trial correct responses from predictive feature, semantic basis, and their congruence (i.e., their interaction, coded as 0.5 when predictive feature matches semantic basis, and -0.5 otherwise). We also included by-participant random intercepts.[1] A likelihood ratio test comparing this full model to reduced models reveals no significant effect of congruence ($\beta_{congruence} = 0.276$; $p = 0.5189$). Our general prediction is thus not confirmed. Additional model comparisons indicate a significant effect of semantic basis ($\beta_{semBasis} = 1.251$, deviation-coding, $p = 0.003$), suggesting participants were better at learning animacy-based noun classes, as in Exp 1.

To test whether congruence facilitates learning for either of the two predictive features separately, we fit two additional logistic models. The first model predicted per-trial correct responses by semantic basis (animacy vs. colour) when animacy was most predictive. The second predicted per-trial correct responses by semantic basis when colour was most predictive. We also included by-participant and by-item[2] random intercepts. Likelihood ratio tests indicate no significant improvement for either model including semantic-basis as a fixed effect (animacy model: $\beta_{semBasis}$ = -0.013, $p = 1$; colour model: $\beta_{semBasis}$ = 0.580, $p = 0.313$; congruent coded as 0.5, incongruent as -0.5). This suggests that, despite a numerical trend when animacy was most predictive, in neither case did congruence facilitate learning.

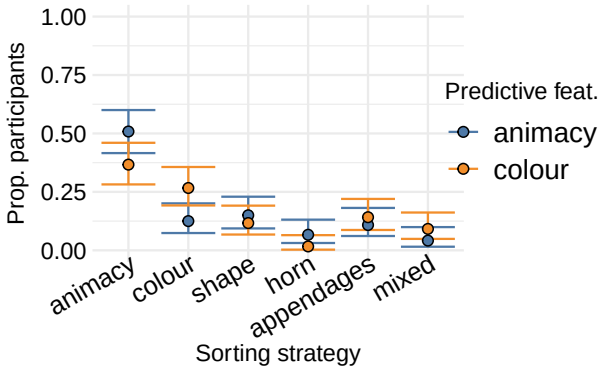

Figure 6: Proportion of Exp 2a participants (y-axis) with each best-fit image sorting strategy (x-axis) by stimulus predictive features (dot colour). *Mixed* aggregates responses with multiple best fits. Error bars are 95% CI around the proportions.

[1] More complex random effects structures resulted in singular fits; we progressively reduced the model and ended up with only the by-participant random intercept.

[2] We fit a by-item random intercept for the colour model only, since including it for the animacy model caused singular fits.

Importantly, not all participants sorted based on the most predictive feature. We therefore compared participants who *did* sort based on the most predictive features—i.e., whose bias was potentially shifted toward the most predictive features—and those who sorted by the least predictive feature—whose bias clearly was not.[3] Numerically, when colour was most predictive, participants who sorted based on colour were better at learning colour-based than animacy-based noun classes (Figure 7, middle panel). However, when animacy was most predictive, no such trend can be observed (Figure 7, left panel). Thus, in Exp 2b, we sought to replicate this with a larger set of participants in the predictive colour conditions.

Exp 2b: Focusing on colour

In Exp 2b we test the prediction that participants who sort by colour when it is most predictive will be better at learning colour-based noun classes than animacy-based ones.

Materials

We used the subset of image stimuli from Exp 2a where colour is more predictive. The artificial language is identical to Exp 2a.

Methods

Participants were randomly assigned to the congruent condition in which the semantic basis of class is colour, or the incongruent condition where it is animacy. The procedure was identical to Exp 2a except that participants who did not sort solely by colour in the sorting task were taken directly to the post-experiment questionnaire and did not take part in the learning phase.

Participants

Participants were native English speakers. We excluded all participants who did not sort the images by colour, or scored less than 90% on the attention check trials or the determine train trials. We kept collecting data until we reached 120 participants after exclusion (60 per condition).

Results

Figure 7 (right panel, green dots) shows the results of Exp 2b, which suggest no difference between animacy and colour-based noun classes. Indeed, a logistic model predicting response accuracy by congruence, with a by-participant intercept[4] revealed no significant effect of congruence ($\beta_{\text{semBasis}} = -0.510$, deviation-coding, $p = 0.314$). In other words, the numerical trend we saw above was not borne out in this larger sample

However, it is clear from the results so far that participants have a prior bias for animacy across these tasks. One possibility is therefore that the manipulation did shift participants' biases, but not strongly enough to *reverse* this bias for animacy-based class systems. In other words, learning a colour-based noun class system is harder, regardless of feature predictiveness. If this is the case, instead of seeing a facilitatory effect of colour predictive power on learning colour-based classes, we might instead see that learning of *animacy-based classes* is hindered. In other words, participants who sort based on colour when it is more predictive might find animacy-based system harder to learn than those who continue to sort based on animacy.

We explored this possibility by comparing participants in Exp 2b, who were exposed to colour predictive stimuli and sorted by colour, to participants in Exp 2a who sorted the same set of stimuli by animacy (N = 44, 22 for each semantic basis). As Figure 7 (right panel) shows, while colour noun class systems were always harder to learn, participants learned *animacy-based* noun classes worse when their bias was shifted toward colour (i.e., they sorted by colour) than when their bias was not shifted toward colour (i.e., they sorted by animacy). We fit a logistic model predicting per-trial correct responses from semantic basis for noun classes, sorting strategy (animacy or colour) and their interaction with a by-participant random intercept. Likelihood ratio tests revealed significant main effects of all three predictors. The significance of semantic basis ($\beta_{\text{semBasis}} = 1.957, p < 0.001$, colour coded as -0.5, animacy as 0.5) confirms that there is an overall bias for animacy. The significance of sorting strategy suggests that participants sorting by animacy (i.e. those from Exp 2a) generally scored higher ($\beta_{\text{sortStrategy}} = -1.350, p = 0.014$, colour coded as 0.5, animacy as -0.5). Crucially, the interaction between the semantic basis for noun classes and sorting strategy was significant ($\beta_{\text{semBasis}\times\text{Sorting}} = 1.440, p = 0.006$, coded as 0.5 when semantic basis matches sorting strategy, -0.5 otherwise) , suggesting that lower predictive power of animacy did weaken the animacy bias. A logistic model predicting per-trial correct responses from sorting strategy including only participants who learned animacy-based noun classes likewise confirms that performance is higher among participants who sorted by animacy ($\beta_{\text{sortStrategy}} = -3.333, p = 0.002$).

General Discussion

The experiments reported here investigate whether the apparent prevalence of animacy-based noun classification and absence of colour-based noun classification is driven by a cognitive bias for the former, and if so what underlies this bias. In Exp 1, we tested whether participants learned animacy-based noun classes more easily than colour-based ones. Our results suggest that indeed they are, providing evidence that learners' biases align with this language universal.

In Exp 2a and 2b, we tested the hypothesis that this bias might be driven by a domain-general principle of categorisation: features higher in predictive power are a better basis of categorisation. In the real-world, animacy is likely to be a better predictor of the other features of an object than colour. We therefore tested our hypothesis by manipulating the predictive power of animacy and colour features in a set

[3] We ignore participants who sorted on other features.

[4] More complex random effects structures caused singular fits.

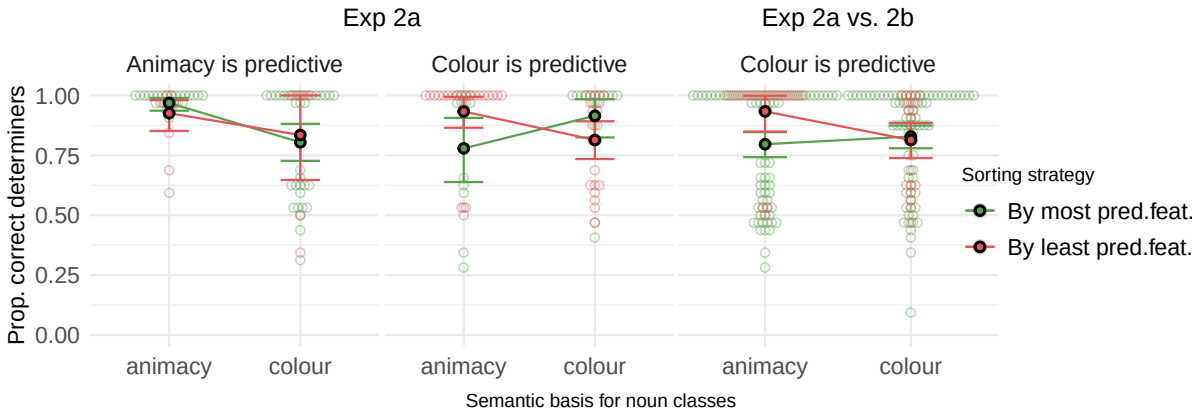

Figure 7: Proportion of correctly selected determiners (y-axis) per participant (unfilled dots) by predictive features (panel), semantic basis (colour), and whether sorting matches the most or the least predictive features (x-axis). Error bars are 95% CI around the means (filled dots). The left and middle panel shows results from Exp 2a. The right panel compares participants under predictive colour conditions across Exp 2a (red, sorting by animacy) and 2b (green, sorting by colour).

of artificially-created unfamiliar objects. We then embedded these in a noun class learning task to see whether predictive power would modulate learning of animacy and colour-based classes. Our results do not straightforwardly confirm this prediction. Across Exp 2a and 2b, we were not able to show that learning was facilitated when the more predictive feature served as the basis of classification. This is likely due to the strong bias for animacy (in both noun class learning and image sorting) combined with the fact that participants did not always notice (i.e., sort based on) the most predictive feature of the stimuli in our task. However, we did find some indication in Exp 2a that when participants actually *did* notice that colour or animacy was more predictive (as evidenced by their image sorting behaviour), they did learn classes based on these features better. In addition, we found that participants shown stimuli in which colour was most predictive were worse at learning animacy-based noun classes when their bias was successfully shifted toward colour, compared to when it was not.

These results constitute the first experimental evidence for the role of predictive power of features in noun classification. This supports the claim that domain-general principles of categorisation may play a role in explaining the pervasiveness of animacy and the absence of colour in noun classification systems. At the same time, it is notable that in both a non-linguistic sorting task and a noun class learning task, participants prefer animacy over colour regardless of our manipulation of predictive features. This could reflect participants prior experience with these features, or something else. Despite this, participants were consistently successful in learning colour-based noun classes across experiments (i.e., they scored consistently around 80%). Whatever underlies the dominance of animacy in these tasks, then, there is no strong bias *against* colour-based classes as one might expect under a theory of hard constraints (e.g., Adger, 2018; Cinque, 2013)

The hypothesis put forward here has implications for theories of language change. In particular, there is a cross-linguistically common historical pathway for nouns to develop into various noun classification systems with progressively fewer categories, i.e., nouns (numerous) $\rightarrow$ classifiers (up to hundreds of classes) $\rightarrow$ noun classes and genders (two to a dozen classes) (Corbett, 1991; Seifart, 2010). If predictive power does indeed shape noun classification, these systems should evolve in a way that preserves the highly, if not maximally, predictive features, given the number of categories at each stage. There is some informal evidence that is the case. Documented cases of the development from noun to classifier appear to retain animacy features, if any feature is retained at all. For example, in Vietnamese, the classifier for inanimates *cái* developed from the noun meaning 'piece' (Löbel, 2000). Similarly, in Akatek, the classifier for humans *winaj* developed from the noun for 'man' (Zavala, 2000). Aikhenvald (2000, p. 401) notes many other similar cases.

Of course, the relationship between our results and this language universal crucially depends on the assumption that animacy-based classifications yield better, i.e., more coherent and distinct, clusters than than colour-based ones. While intuitively correct, this should also be empirically verified, for instance, by comparing the quality of animacy- and colour-based clusterings of word embeddings (Mikolov, Chen, Corrado, & Dean, 2013).

Finally, as noted in the introduction, our study contributes to a growing body of empirical research that highlights the role of cognitive biases in shaping language. Since high predictive power is entailed by maximally compact categorisation, the bias for animacy over colour in noun classification may ultimately be an instantiation of a highly general simplicity bias, which accounts for not only universals in language (Culbertson & Kirby, 2016) but also phenomena across cognitive domains (Chater & Vitányi, 2003).

Acknowledgments

The authors would like to thank anonymous reviewers for helpful comments and constructive feedback. This project was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 757643). Ponrawee Prasertsom is grateful for support for his studies from the Anandamahidol Foundation, Thailand.

References

Adger, D. (2018). The autonomy of syntax. In N. Hornstein, H. Lasnik, P. Patel-Grosz, & C. Yang (Eds.), *Syntactic structures after 60 years* (pp. 153–176). De Gruyter. doi: 10.1515/9781501506925-157

Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices*. OUP Oxford.

Aikhenvald, A. Y. (2017). A typology of noun categorization devices. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *The cambridge handbook of linguistic typology* (p. 361–404). Cambridge University Press.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological review*, *98*(3), 409.

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in cognitive sciences*, *7*(1), 19–22.

Cinque, G. (2013). Cognition, universal grammar, and typological generalizations. *Lingua*, *130*, 50-65. (SI: Syntax and cognition: core ideas and results in syntax) doi: 10.1016/j.lingua.2012.10.007

Corbett, G. G. (1991). *Gender*. Cambridge University Press.

Corbett, G. G. (2013). Sex-based and non-sex-based gender systems (v2020.3) [Data set]. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.7385533 doi: 10.5281/zenodo.7385533

Croft, W. (2002). *Typology and universals*. Cambridge University Press.

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in psychology*, *6*, 1964.

Culbertson, J., & Kirby, S. (2022). Syntactic harmony arises from a domain-general learning bias. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, *139*, 71–82.

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351. Retrieved from https://doi.org/10.21105/joss.05351 doi: 10.21105/joss.05351

D'Alessandro, R. (2021). Not everything is a theory. *Theoretical Linguistics*, *47*(1-2), 53–60. doi: 10.1515/tl-2021-2005

Fedzechkina, M., & Jaeger, T. F. (2020). Production efficiency can cause grammatical change: Learners deviate from the input to better balance efficiency against robust message transmission. *Cognition*, *196*, 104-115. doi: 10.1016/j.cognition.2019.104115

Griffiths, T. L., Sanborn, A. N., Canini, K. R., Navarro, D. J., & Tenenbaum, J. B. (2011). Nonparametric bayesian models of categorization. *Formal approaches in categorization*, 173–198.

Holmes, K. J., & Regier, T. (2017). Categorical perception beyond the basic level: The case of warm and cool colors. *Cognitive science*, *41*(4), 1135–1147.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049-1054. doi: 10.1126/science.1218811

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

Kramer, R. (2020). Grammatical gender: A close look at gender assignment across languages. *Annual Review of linguistics*, *6*, 45–66.

Lee, M. (1988). Language, perception and the world. In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 211–246). Blackwell.

Löbel, E. (2000). Classifiers versus genders and noun classes: A case study in Vietnamese. *Trends in linguistics studies and monographs*, *124*, 259–320.

Maldonado, M., Zaslavsky, N., & Culbertson, J. (2023). Evidence for a language-independent conceptual representation of pronominal referents. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space.*

Newmeyer, F. J. (2005). *Possible and probable languages: A generative perspective on linguistic typology*. Oxford University Press, USA.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive science*, *26*(3), 303–343.

Pothos, E. M., Chater, N., & Hines, P. (2011). The simplicity model of unsupervised categorization. *Formal approaches in categorization*, 199–219.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. doi: 10.1016/0010-0285(76)90013-X

Seifart, F. (2010). Nominal classification. *Language and Linguistics Compass*, *4*(8), 719–736.

Talmy, L. (1988). The relation of grammar to cognition. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics.* John Benjamins Publishing Company.

Vinh, N. X., Epps, J., & Bailey, J. (2010). In-

formation theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, *11*(95), 2837–2854. Retrieved from http://jmlr.org/papers/v11/vinh10a.html

Zavala, R. (2000). Multiple systems of classifiers in akateko. In *Systems of nominal classification* (p. 114–146). Berlin, New York: Mouton de Gruyter.