In search of a unified mechanism for regularisation of linguistic variation

Aislinn Keogh, Simon Kirby and Jennifer Culbertson

Centre for Language Evolution, University of Edinburgh
aislinn.k.keogh@gmail.com

1 Introduction

Natural languages rarely exhibit variation that is not predictable according to some criteria [7]. For example, there is considerable variation in English plural marking, but the choice of variant is lexically-conditioned – so *dog* → *dogs*, but *goose* → *geese* and *fish* → *fish*. Experimental and computational work (e.g. [6, 20, 11, 17]) has demonstrated that individuals have cognitive biases which work against *unpredictable* variation. When no conditioning is present, learners tend to regularise variation (e.g. by increasing use of one rule to the exclusion of others).

Furthermore, even *predictable* variation is sometimes lost as languages evolve, with previously irregular forms (e.g. *cow* → *kine* in Middle English) adopting the regular pattern. Does this kind of language change reveal a similar bias against predictable variation? The most parsimonious explanation is that predictable and unpredictable variation are not two fundamentally different phenomena, but that they exist on one spectrum of complexity, with greater predictability equating to greater simplicity and the simplest possible system being one with no variation at all [3]. If this is the case, the same cognitive biases that act on unpredictable variation should also reduce even predictable variation, albeit potentially less strongly or over a longer time frame. However, there is some experimental evidence that conditioned variation is treated differently from unconditioned variation [9, 11, 18]. Pragmatics-based accounts of regularisation (e.g. [14]) also predict that, while languages will not sustain unpredictable variation, conditioned variation should be evolutionarily stable.

In the present study, we use an artificial language learning experiment to investigate whether the same cognitive biases target both predictable and unpredictable variation. Experimental evidence consistently points to a production-side bias for regularity rather than a learning-based account [6, 14, 11]. More specifically, limitations on memory retrieval during language production have been suggested as one potential mechanism for the loss of *unpredictable* variation [10]. Here, we further test this hypothesis, providing evidence that regularisation of both predictable and unpredictable variation arise via this same mechanism.

We train participants on a language exhibiting either predictable or unpredictable variation in plural marking, and with working memory taxed during learning, production, or not at all. In line with the production-side account of regularisation, we predicted that participants across conditions would show no evidence of having encoded a more regular language than the one they were trained on. In line with the memory retrieval hypothesis, we predicted that we would see the clearest evidence for reduction of variation when taxing working memory during production. Finally, to test our hypothesis regarding variation type, we predicted that the effect of memory limitations during production would be modulated by variation type, with greater regularisation of unpredictable languages.

2 Methods

The study was pre-registered with the Open Science Foundation (https://osf.io/vqyej).

Design: Participants were randomly assigned to one of six conditions in a 2x3 between-subjects design. Their input languages exhibited either probabilistically lexically-conditioned (PREDICTABLE conditions) or random (UNPREDICTABLE conditions) variation. To create a memory pressure, we use an interference task known to specifically tax working memory [13]. This was administered either during training (LEARNING LOAD conditions), production (PRODUCTION LOAD conditions) or not at all (NO LOAD conditions).

Participants: Participants were 220 adult, self-reported native English speakers with no known language disorders, recruited via Prolific and paid £3 for up to 20 minutes' participation. Following our pre-registered exclusion criteria, we excluded 47 participants from analysis for the following reasons: self-reporting the use of written notes in an exit questionnaire contrary to instructions (3), data saving errors (1), failing to provide usable data on more than two critical trials (38), and button mashing (5). This left us with data from 173 participants[1].

Materials: Participants were asked to learn a small artificial language consisting of text labels paired with six images. Each image depicted a pair of animals, and was described by a two-word label: one word for the noun and one word indicating plurality (presented in the English frame 'Here are two...'). Noun labels were iconic (e.g. "buzzo" for a bee) and paired with one of two plural markers, both non-English CVC monosyllables ("mej" and "huv"). The mapping of nouns to plural markers varied according to condition as shown in Table 1.

	Predictable							**Unpredictable**						
	N1	**N2**	**N3**	**N4**	**N5**	**N6**	**Total**	**N1**	**N2**	**N3**	**N4**	**N5**	**N6**	**Total**
P1	7	7	7	7	1	1	30	5	5	5	5	5	5	30
P2	1	1	1	1	7	7	18	3	3	3	3	3	3	18
Total	8	8	8	8	8	8	48	8	8	8	8	8	8	48

Table 1: Distribution of plural markers (P_i) across nouns (N_j) in the two variation conditions. In PREDICTABLE conditions, four nouns were randomly assigned to one plural marker (the 'regulars') and two to the other marker (the 'irregulars'). A small amount of noise was then added to this mapping, such that each noun appeared with its assigned plural marker 87.5% of the time, and with the other 12.5% of the time. This noisy conditioning meant that participants could regularise without having to produce a description they had never observed. In UNPREDICTABLE conditions, all nouns appeared with one marker 62.5% of the time, and with the other 37.5% of the time. Both markers appeared with the same overall frequency in the two variation conditions. The assignment of specific nouns to the N slots and specific plural markers to the P slots was randomised for each participant.

Procedure: The experiment was written in JavaScript using the JsPsych library [5] and ran in participants' web browsers. Participants were randomly assigned to one of the six conditions at the start of the experiment. In all conditions, the experiment consisted of three phases: training, production, and estimation. The order of presentation was randomised within each phase. A schematic of the experiment is given in Figure 1.

On each training trial, participants were shown an image and a corresponding description. Each of the six images was shown eight times for a total of 48 trials. On each production trial, participants saw an image and had to complete the description by clicking two buttons from an array consisting of all nouns and all plural markers in the language. As in training, each of

[1] PREDICTABLE/NO LOAD: 29; PREDICTABLE/LEARNING LOAD: 30; PREDICTABLE/PRODUCTION LOAD: 29; UNPREDICTABLE/NO LOAD: 28; UNPREDICTABLE/LEARNING LOAD: 28; UNPREDICTABLE/PRODUCTION LOAD: 29

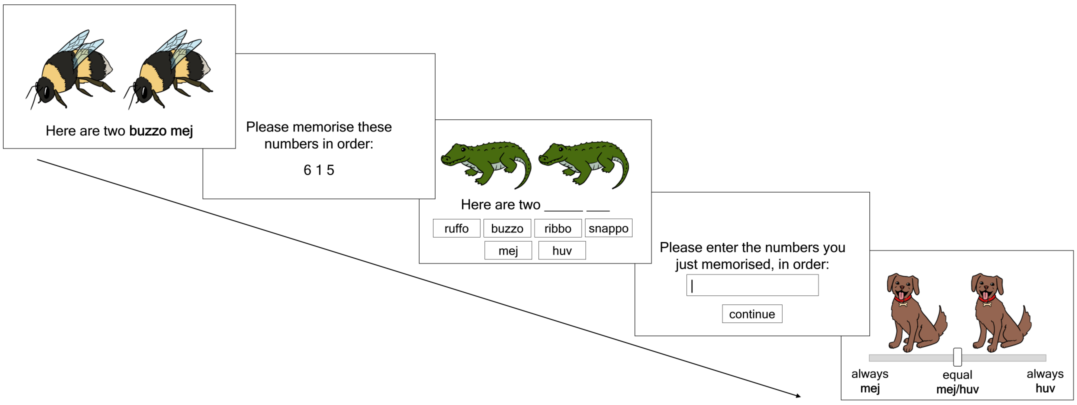

Figure 1: Schematic of the experiment: PRODUCTION LOAD condition. Top to bottom: training trial, digit sequence presentation, production trial, digit sequence recall, estimation trial. Participants in LEARNING LOAD conditions would instead have seen the digit sequence presentation and recall trials sandwiched around each training trial. Participants in NO LOAD conditions would not have seen these digit sequence trials.

the six images was shown eight times for a total of 48 trials. Finally, in the estimation phase, participants were shown all six images and asked to estimate how often they had seen each noun with each plural marker in training. They provided these estimates using a continuous slider with three labels: "always *plural 1*" on the left, "equal *plural 1/plural 2*" in the middle, and "always *plural 2*" on the right.

In LEARNING LOAD and PRODUCTION LOAD conditions, participants additionally completed a digit sequence recall task (modelled after [13]) during the training or production phase, respectively. Before every trial, participants were asked to memorise a pseudo-random sequence of three digits. Immediately after each trial, they were asked to retype the digits. They were given feedback on how many digits they had recalled in the correct position and how long they had taken to respond, to encourage both speed and accuracy.

3 Results

Following previous work [6, 14, 20], we quantify regularisation in information theoretic terms [19]. Here, we measure the entropy of the plural marking system participants estimated and produced, relative to the input language. When entropy decreases, we can infer that one variant has become more frequent overall. The input languages are matched on this measure (0.95 bits). The degree of lexical conditioning – or predictability – of plural marking is given by the mutual information between nouns and plural markers. Again, we calculate this for participants' estimates and productions, relative to the input. When mutual information increases, we can infer that variation has become more strongly lexically-conditioned. Unlike for entropy, the input languages differ in mutual information: UNPREDICTABLE languages score 0 (indicating no lexical conditioning), while PREDICTABLE languages score 0.41 (indicating imperfect lexical conditioning). Our hypothesis is that regularisation of both types of variation arises from a production-side bias driven by memory limitations. If this is the case, we would expect no reliable change in either measure between the input and participants' *estimates*, but a reliable change in at least one of the measures between the input and participants' *productions* in PRODUCTION LOAD conditions, especially for UNPREDICTABLE languages (which are more complex [3]).[2]

We first investigate regularisation during learning by analysing the change in entropy between participants' training data and the languages described by their estimates, as shown in Figure 2a. We found no reliable decrease in entropy: confidence intervals around the mean for all conditions are either above or crossing zero.

However, as Figure 2b suggests, there was an increase in mutual information between the languages participants in UNPREDICTABLE conditions were trained on and the ones described by their estimates: confidence intervals around the mean are all above zero for these conditions. Pairwise comparisons between conditions further reveal that, while there are no differences between memory load conditions within each variation type, every UNPREDICTABLE condition is reliably different from every PREDICTABLE condition.

To summarise, our results suggest that while overall variation is not reduced during learning, learners *are* biased to infer patterns of conditioning when variation in the input is unpredictable.

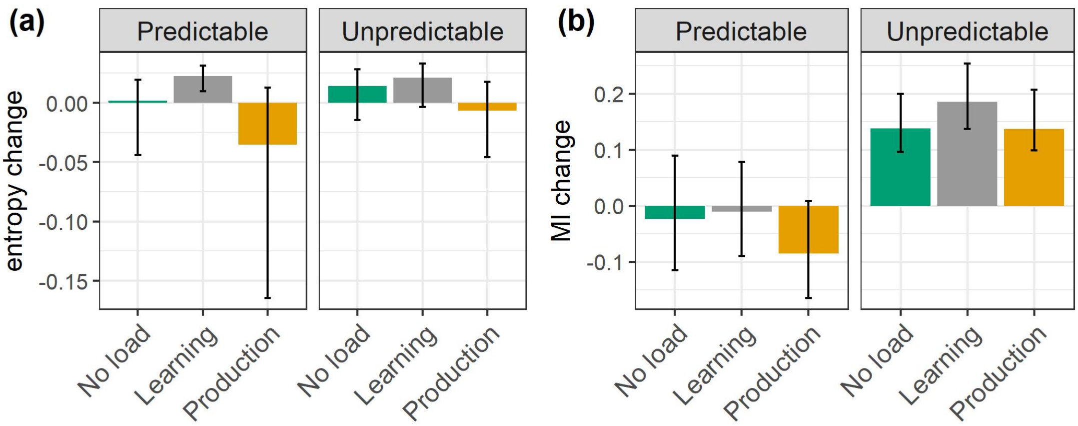

Figure 2: Change in entropy (left) and mutual information (right) from the training data to participants' estimates, by condition. Error bars represent bootstrapped 95% confidence intervals over the mean.

We next investigate regularisation during production by analysing the change in entropy between participants' training data and the languages they produced, as shown in Figure 3a. Here we find evidence for a drop in entropy in PRODUCTION LOAD conditions only (confidence intervals around the mean are below zero for both these conditions, but cross zero for all other conditions). However, we find no evidence that this effect is stronger in the UNPREDICTABLE condition[3].

As suggested by Figure 3b, we observed an increase in mutual information across all conditions except PREDICTABLE/PRODUCTION LOAD: confidence intervals around the mean are

[2]Inspection of the models specified in our pre-registration revealed that residuals were significantly non-normally distributed and had non-constant variance over groups. Since our data did not meet the assumptions for a linear modelling analysis, the analyses we present here instead evaluate our pre-registered predictions using 95% bootstrapped confidence intervals around the mean of each condition, as well as around the differences between condition means (following [12]). For each analysis, we used R [16] to calculate bias-corrected and accelerated (BCa) bootstrap intervals (as recommended by [15]) from 10,000 samples. The pattern of results from this analysis is identical to the one obtained from linear models.

[3]We tested for this interaction by generating, for each variation type, a set of bootstrap estimates of the difference between PRODUCTION LOAD and other memory load conditions (collapsed). The 95% bootstrapped confidence interval around the difference between these differences crosses zero, indicating no interaction.

above zero in all but the latter. On this measure, our data therefore suggest that there is a general bias towards introducing or boosting lexical conditioning, not arising from memory limitations during *production*. In fact, pairwise comparisons between conditions indicate that mutual information increased more in the UNPREDICTABLE/LEARNING LOAD condition than all others, suggesting that the preference for lexically-conditioned patterns of variation is actually amplified by memory limitations during *learning*.

To summarise, our results suggest that reduction in overall variation is driven by memory limitations during *production*. By contrast, a bias for lexical conditioning is found almost across the board, and is, if anything, more pronounced when memory is taxed during *learning* of unpredictable variation.

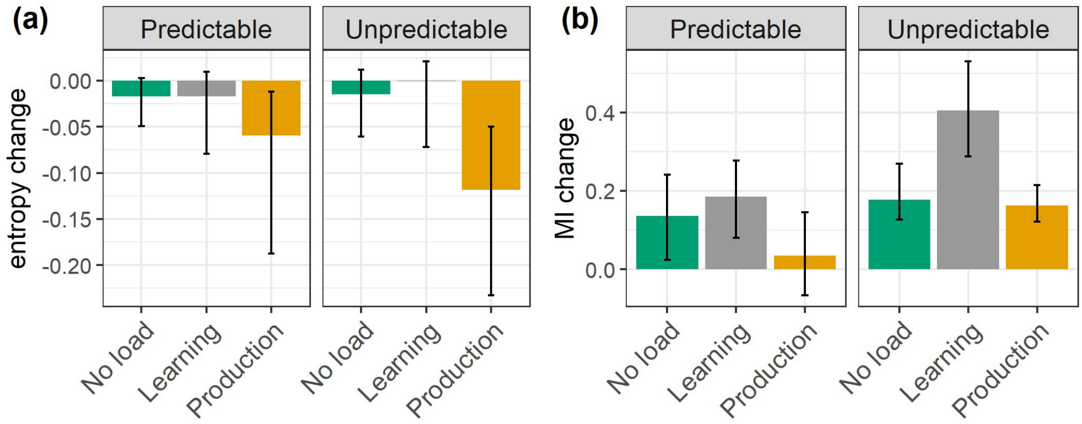

Figure 3: Change in entropy (left) and mutual information (right) from the training data to participants' productions, by condition. Error bars represent bootstrapped 95% confidence intervals over the mean.

4 Discussion

In this experiment, we investigated whether similar cognitive mechanisms drive regularisation of predictable (conditioned) and unpredictable (random) variation. Both types of variation exist in language, but there is historical evidence that both are lost over time, and experimental evidence that at least unpredictable variation is regularised – potentially due to memory limitations during production. Our findings partially align with previous research (e.g. [6, 11]) suggesting that regularisation is not driven by learners' failure to accurately encode the total amount of variation in their input. In particular, we found evidence for a reduction in overall variation during production, but not learning. This effect was indeed driven by memory load. However, our results point to a clear bias in favour of predictable patterns of variation, as evidenced by the increase in lexical conditioning during both learning and production. This effect was further amplified by memory limitations operating during learning, but is evident even in the absence of any memory load. In other words, a bias to reduce variability by increasing conditioning affects both language users' inferences during learning and their (implicit) decisions during production.

We began with the question of whether predictable and unpredictable variation are two points on the same spectrum. While our results suggest that some of the same mechanisms target both, we also find evidence that the introduction of lexical conditioning is not simply a step on the way to eliminating variation altogether: the latter process appears to happen exclusively during production, while the increase in predictability is at least partially accounted for by learning effects. Furthermore, in both learning and production, we see much bigger changes in mutual information than in entropy, suggesting that the bias in favour of predictability is much stronger than the bias in favour of complete uniformity. A question for future work is how the relative strength of these biases interacts with the number of variants: with only two plural markers, as in our design, it is presumably not difficult to maintain both. Expanding the set of variants may heighten the pressure to reduce variation (although see [11]). A different kind of production task – with participants required to free-type their descriptions rather than being cued by having the whole lexicon presented as buttons – would also make retrieval more taxing [10]. We would predict that both of these modifications would result in greater entropy drop, especially under memory load during production.

It is also worth noting that, although our analysis shows a reliable reduction in entropy in the PRODUCTION LOAD conditions overall, there is considerable variation in how different participants respond to the memory manipulation. An open question is whether it is enough for only some individuals in a language community to have very strong biases against variation for that variation to be lost through cultural evolution. Certainly, the substantial individual differences we see here suggest at least that we should expect diachronic loss of variation from natural languages to be a slow process.

At a population level, though, our results suggest that language users have biases which work against variation of all kinds. From the perspective of language evolution, one might therefore wonder why variation of the kind described here is so pervasive in natural languages. As for any cognitive bias shaping language, the explanation for this is likely a combination of the fact that these biases are weak (i.e. defeasible), and compete with other pressures shaping language. Usage-based accounts (e.g. [1]) argue that systems of conditioned variation are maintained because the easiest variant to access in any given context is simply the one that has been experienced most often in that context (rather than the one that has been experienced most often overall). Furthermore, although this was not relevant in our task, this kind of variation may persist due to frequency-dependent patterns in regularity: irregular forms tend to be highly frequent, presumably making it easier to learn and retrieve the correct form [4].

5 Conclusion

We have provided evidence in favor of the claim that cognitive biases leading to regularisation target both unpredictable and predictable variation. Our findings support the idea that regularisation is particularly strong during production, and is driven at least in part by memory limitations. However, our results also suggest – contrary to our predictions and at least some previous work – that this memory bottleneck may play a distinct role during both learning and production. Specifically, while over-retrieval of the most accessible variant [8] during language production may act to reduce overall variability, *unpredictability* appears to decrease more as a result of inferences formed during learning. Overall, the results of this study lend support to the notion that cognitive constraints in individuals can give rise to particular structures in languages. By allowing languages to pass more easily through the bottleneck imposed by working memory limitations [2], we argue that properties like regularity are more likely to be selected through cultural evolution and thus develop into typological universals.

References

[1] Joan Bybee. From usage to grammar: The mind's response to repetition. *Language*, 82(4):711–733, 2006.

[2] Morten H. Christiansen and Nick Chater. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, 2016.

[3] Jennifer Culbertson and Simon Kirby. Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, 6:1964, 2016.

[4] Christine Cuskley, Martina Pugliese, Claudio Castellano, Francesca Colaiori, Vittorio Loreto, and Francesca Tria. Internal and external dynamics in language: Evidence from verb regularity in a historical corpus of English. *PLOS ONE*, 9(8):e102882, 2014.

[5] Joshua R. de Leeuw. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1):1–12, 2015.

[6] Vanessa Ferdinand, Simon Kirby, and Kenny Smith. The cognitive roots of regularization in language. *Cognition*, 184:53–68, 2019.

[7] Talmy Givón. Function, Structure and Language Acquisition. In Dan Isaac Slobin, editor, *The Crosslinguistic Study of Language Acquisition*, pages 1005–1028. Lawrence Erlbaum Associates Inc., 1985.

[8] Adele Goldberg and Fernanda Ferreira. Good-enough language production. *Trends in Cognitive Sciences*, 26(4):300–311, 2022.

[9] Carla L. Hudson Kam. The impact of conditioning variables on the acquisition of variation in adult and child learners. *Language*, 91:906–937, 2015.

[10] Carla L. Hudson Kam and Ann Chang. Investigating the cause of language regularization in adults: Memory constraints or learning effects? *Journal of Experimental Psychology: Learning Memory and Cognition*, 35(3):815–821, 2009.

[11] Carla L. Hudson Kam and Elissa L. Newport. Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, 59(1):30–66, 2009.

[12] Yasamin Motamedi, Lucie Wolters, Danielle Naegeli, Simon Kirby, and Marieke Schouwstra. From improvisation to learning: How naturalness and systematicity shape language evolution. *Cognition*, 228:105206, 2022.

[13] Andrew Perfors. When do memory limitations lead to regularization? An experimental and computational investigation. *Journal of Memory and Language*, 67(4):486–506, 2012.

[14] Andrew Perfors. Adult regularization of inconsistent input depends on pragmatic factors. *Language Learning and Development*, 12(2):138–155, 2016.

[15] Marie-Therese Puth, Markus Neuhäuser, and Graeme D. Ruxton. On the variety of methods for calculating confidence intervals by bootstrapping. *Journal of Animal Ecology*, 84(4):892–897, 2015.

[16] R Core Team. R: A Language and Environment for Statistical Computing, 2020.

[17] Florencia Reali and Thomas L. Griffiths. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3):317–328, 2009.

[18] Anna Samara, Kenny Smith, Helen Brown, and Elizabeth Wonnacott. Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, 94:85–114, 2017.

[19] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[20] Kenny Smith and Elizabeth Wonnacott. Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3):444–449, 2010.