THE UNIVERSITY OF
WARWICK

Original citation:
Smith, Kenny and Wonnacott, Elizabeth. (2010) Eliminating unpredictable variation through iterated learning. Cognition, Vol.116 (No.3). pp. 444-449.

Permanent WRAP url:
http://wrap.warwick.ac.uk/51390

warwick**publications**wrap

highlight your research

http://wrap.warwick.ac.uk/

Eliminating unpredictable variation through iterated learning

Kenny Smith & Elizabeth Wonnacott

Abstract

Human languages may be shaped not only by the (individual psychological) processes of language acquisition, but also by population-level processes arising from repeated language learning and use. One prevalent feature of natural languages is that they avoid unpredictable variation. The current work explores whether linguistic predictability might result from a process of iterated learning in simple diffusion chains of adults. An iterated artificial language learning methodology was used, in which participants were organized into diffusion chains: the first individual in each chain was exposed to an artificial language which exhibited unpredictability in plural marking, and subsequent learners were exposed to the language produced by the previous learner in their chain. Diffusion chains, but not isolate learners, were found to cumulatively increase predictability of plural marking by lexicalising the choice of plural marker. This suggests that such gradual, cumulative population-level processes offer a possible explanation for regularity in language.

1.   Introduction

To what extent are human behaviours a straightforward reflection of the underlying psychological characteristics of the individual? This is a key question in the cognitive sciences, and is central to the debate in linguistics over the relationship between the observed typological distribution of languages and psychological constraints on language acquisition (see e.g. Chomsky, 1965; Christiansen & Chater, 2008; Evans & Levinson, 2009): are the languages we see in the world a reflection of strong or even absolute constraints on possible languages imposed during acquisition, or might they also be a consequence of the interaction of multiple weaker constraints arising from acquisition and use?

To take a specific example: one property of human language is that variation tends to be *predictable*. In general, no two linguistic forms will occur in precisely the same environments and perform precisely the same functions (Givón, 1985). Instead, usage of alternate forms is conditioned in accordance with phonological, semantic, pragmatic or sociolinguistic criteria. Conditioning of variation occurs at all levels of linguistic structure, including phonetics (e.g. sociolinguistic conditioning of vowel variants in English: Labov, 1963), morphology (e.g. phonological conditioning of plural allomorphs in English: Lass, 1984, p. 13-14), and syntax (e.g. semantic conditioning of noun classes in Dyirbal: Dixon, 1972; sociolinguistic and syntactic conditioning of copula/auxiliary BE in Bequia: Meyerhoff, 2008).

Several recent studies have investigated whether this predictability might be a consequence of constraints inherent in language acquisition. One route to address such questions is through the use of *artificial language learning* paradigms, where experimental participants are trained and tested on experimenter-designed miniature languages. One consistent finding from this literature is that, given a language in which two forms are in free variation, adult learners tend to *probability match*, i.e. produce each variant in accordance with its relative frequency in the input, although they may regularize in certain specialized circumstances (Hudson Kam & Newport, 2005, 2009; Wonnacott and Newport, 2005). There is also evidence that children are more likely to regularize than adults (Hudson Kam & Newport, 2005, 2009; Austin, Newport & Wonnacott, 2006), although they may probability match in some circumstances (Wonnacott & Perfors, 2009).

Findings of this nature feed into the debate on the role of adult and child learners in processes of language change and language formation via creolization. Elimination of variation via analogical levelling – a form of regularization – is a key process in language change (see e.g. Hock, 2003), and creolization can also be characterised as the construction of a new language via levelling and regularization of a pool of linguistic variants arising from radical language contact (Siegel, 2004). One possible implication of the differences in adult and child treatment of unpredictable variation, as highlighted by Hudson Kam & Newport (2005), is that child learners may be primarily responsible for the elimination of variability during language change and creolization. However, this conclusion seems at odds with at least some of the literature on language change and creolization, which emphasises the role of adult learner/users (e.g. Croft, 2000; Mather, 2006).

The experimental studies discussed above explore the changes in linguistic systems arising from *individual* processes of acquisition. However, languages may also be shaped by processes which are the product of *populations,* i.e. collections of multiple individuals: populations may exhibit collective behaviours which differ from the behaviours of isolated individuals, as a consequence of individuals in those populations interacting with, and learning from, one another. For example, symbolic and structured communicative behaviours have been shown to arise through (communicative or learning) interactions between adults in laboratory contexts (e.g. Garrod et al., 2007; Kirby, Cornish & Smith, 2008).

Furthermore, the process of *iterated learning* (where learners observe and learn a behaviour which is itself learned) may provide greater insights into the biases of individual learning than can be obtained in individual-based experiments: under certain circumstances, iterated learning amplifies those biases, potentially making weak biases more apparent (Griffiths & Kalish, 2007; Kalish, Griffiths & Lewandowsky, 2007; Kirby, Dowman & Griffiths, 2007; Griffiths, Christian & Kalish, 2008; Reali & Griffiths, 2009). For example, using a similar methodology to that described here, Reali & Griffiths (2009) show that apparently weak learner biases against synonymy are amplified over repeated episodes of learning, so that a lexicon with multiple labels for objects develops into one with unique, predictable object labelling.

In this paper we use a simple model of a population, namely a *diffusion chain* (where the output of one learner forms the input to the next learner in a chain of transmission), in order to explore the impact of cultural transmission on linguistic variability. Even given our rather minimal population model and the limited interaction between individuals it allows, we find that transmission in populations leads to linguistic systems which differ markedly from those of individual learners: specifically, we show that, in circumstances where individual adult learners would preserve unpredictable variation, simple diffusion-chain populations exhibit cumulative regularization as a consequence of iterated learning. We use plural marking as a simple test-case, and initialise a series of diffusion chains with semi-artificial languages which exhibit unpredictable variability in plural marking: two possible plural markers are used interchangeably. The language is then transmitted from learner to learner according to the standard diffusion chain method. The end result of this process is a linguistic system which still exhibits variability, but that variability is *predictable:* choice of plural marker comes to be conditioned on the linguistic context, namely the noun being marked. This has implications for our understanding of the link between the psychology of the individual and the structure of socially-learned behaviours such as language, and therefore speaks directly to processes of language change and creolization.

2. Method

2.1 Participants

65 monolingual English-speaking undergraduate Psychology students at Northumbria University participated in the study, as part of a participation cooperative. 50 of these participants were involved as part of a diffusion chain (see below), the remainder were included as isolated individuals (henceforth *isolates*).

2.2   Procedure

The learning procedure was identical for all participants. Participants worked through a computer program[1] which presented and tested them on a semi-artificial language. The language was text-based: participants observed objects and text displayed on the monitor and entered their responses using the keyboard.

2.2.1   Language Learning and Testing Procedures

Participants progressed through a three-stage training and testing regime:

1) *Noun familiarization:*   Participants viewed pictures of four cartoon animals (cow, pig, giraffe, rabbit) along with English nouns (e.g. "cow" – hence the designation *semi*-artificial). Each presentation lasted 2 seconds, after which the text (but not the picture) disappeared and participants were instructed to retype that text. Participants then viewed each picture a second time, without accompanying text, and were asked to provide the appropriate label via typing.

2) *Sentence learning:* Participants were exposed to sentences (drawn either from the experimenter-designed input language, for isolates and the first participant in each diffusion chain, or the language generated by the previous learner in their diffusion chain: see below) paired with visual scenes. Scenes showed either single animals or pairs of animals (of the same type) performing a "move" action, depicted graphically using an arrow. Sentences were presented in the same manner as nouns (participants viewed a visual scene plus text, then retyped the text). Each of the eight scenes was presented 12 times (12 training blocks, each block containing one presentation of each scene, order randomized within blocks).

3) *Sentence testing:* Participants viewed the same eight scenes without accompanying text and were asked to enter the appropriate sentence. Each of the eight scenes was presented four times (four blocks, order randomized within blocks).

2.2.2   Initial Input Language

The following language was used with isolates and the first participant in each diffusion chain:

Vocabulary:
Nouns: *cow, pig, giraffe, rabbit*
Verb: *glim* ("move")
Plural markers: *fip, tay*

Sentences:
All sentences were of the following form:
glim NOUN (singular NOUN moves; e.g. *glim cow* = cow moves)
glim NOUN fip/tay (plural NOUN moves; e.g. *glim cow tay* = cows move)

---
[1] Developed using Slide Generator: http://www.psy.plymouth.ac.uk/research/mtucker/SlideGenerator.htm.

The critical feature of the input language was the usage of *fip* and *tay.* One marker was three times more frequent than the other: 5 chain-initial participants and 8 isolates were presented with a language where 75% of plurals were marked with *fip* and 25% of plurals were marked with *tay*, 5 chain-initial participants and 7 isolates were presented with the complement language (25% *fip*, 75% *tay*). Importantly, these statistics also applied to each noun: each noun was paired with the more frequent plural marker nine times and the less frequent marker three times during training. Plural marking in the input language is therefore unpredictable: while one marker is more prevalent, both markers occur with all nouns.

2.2.3   Diffusion Chain Design

50 participants were organised into ten diffusion chains[2] of five individuals, with the initial participant in each chain being trained on the input language specified above and each subsequent individual in a given chain being trained on the language produced during testing by the preceding participant in that chain (with each testing block forming the basis for three training blocks). To convert test output from participant *n* into training input for participant *n+1*, for a given scene, we simply inspected whether participant *n* used *fip*, *tay*, or no marker, and used this marking when training participant *n+1*. In situations where the marker was mistyped, we treated it as if the participant had produced the closest marker to the typed string, based on string edit distance (e.g. "tip" treated as *fip*). Errors in the verb or noun used were not passed on to the next participant, in order to focus on the variability of the language along a single well-defined dimension.[3] Each test block from participant *n* was reduplicated to generate 3 training blocks for participant *n+1*: the order of participant *n*'s four test blocks was randomized, then that sequence of four blocks was presented three times in succession during training of participant *n+1*.[4]

3.   Results

3.1   Number of markers produced

Figure 1 shows the number of plurals[5] marked with the chain-initial majority marker (i.e. *fip* for chains initialised with 75% *fip* marking) for each participant in all 10 chains.[6] A repeated

---

[2] See Mesoudi & Whiten (2008) and Whiten & Mesoudi (2008) for reviews of the diffusion chain method.

[3] Of 2080 sentences entered during testing, participants produced sentences with word order *glim NOUN (particle)* 2069 times (10 of the 11 non-conforming sentences omitted the verb). On eight occasions the wrong noun was provided. On 86 occasions the verb was mis-typed (most common error "gilm", 73 occurrences). On 22 occasions a marker other than null/*fip*/*tay* was used. Eight of those errors were "flip" (corrected to *fip*). The next most common was "fay", corrected to *tay*, six occurrences.

[4] We ran a second experiment (N=40, organised into eight diffusion chains), identical in all respects to the experiment described here but where each participant completed 12 test blocks, rather than 4, with each test block for participant *n* providing a single training block for participant *n+1*, order of blocks randomized. This second experiment replicates the results described here.

[5] Only sentences describing two-animal scenes are considered in the analyses here.

[6] No effect was found for input language (majority *fip* versus majority *tay*) on either proportion of marking or conditional entropy (chains: $F(1,8) \leq 1.705$, $p \geq 0.228$; isolates only: $t(13) \leq 0.656$, $p \geq 0.523$; all individuals exposed to input language [i.e. isolates plus first participant in each chain]: $t(23) \leq 0.989$, $p \geq 0.333$). Consequently, we collapsed results across both input languages.

measures ANOVA reveals no effect for position in chain on the proportion of plurals marked with the majority marker ($F_{(2.746, 21.972)}=1.335$, $p=0.288$, Huynh-Feldt correction). This is consistent with probability-matching behaviour: participants copy the proportion of marking that they see, and that proportion of marking is (on average) preserved across all five participants in a chain. A more powerful test combining the ten chain-initial participants with the 15 isolates also suggests that learners exposed to an unpredictable language reproduce approximately the same distribution of markers that they received in their data (one-sample t-test against the 12 uses of the majority marker, mean difference = 0.84 fewer uses of that marker, SD=2.46, $t_{(24)}=-1.707$, $p=0.101$).
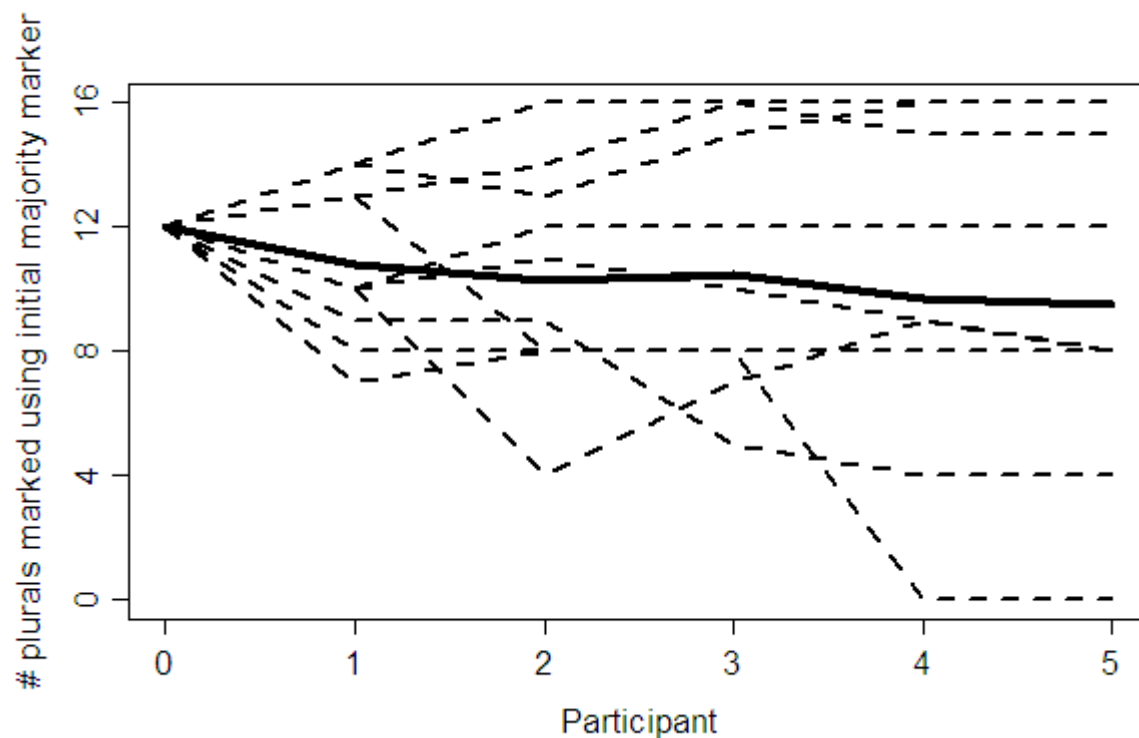

Figure 1: Number of plurals marked using the marker which was initially in the majority in each chain (out of 16 two-animal scenes encountered by each participant during testing). Solid line gives mean, dashed lines show individual chains. Participant 0 is the experimenter-designed input language used to train the first participant in each chain.

3.2 Predictability of marker use
The analyses above are consistent with an account in which individual adults probability match. However, as illustrated in Figure 1, the separate diffusion chains diverge over time: while the proportion of marking across chains matches that of the input languages, individual chains converge towards a range of end points, namely 100%, 75%, 50%, 25% or 0% marking with the majority initial marker. This divergence is unexpected under a pure probability-matching account.

The proportion of markers used does not attest to the predictability of their usage: within a given proportion of plural marking, both predictable and unpredictable systems are possible. For example, in the initial 75%-25% languages there was (by design) no consistent relationship between the noun being marked and the choice of marker: each noun occurred

nine times with one marker and three times with the other. However, we can imagine another 75%-25% language where the choice of marker is entirely predictable: for instance, *tay* might always be used with *giraffe*, with *fip* being used with all other nouns. We can capture this notion of predictability by measuring the *conditional entropy* (*H*) of markers given the noun being marked:

$$H = - \sum_{n \in N, m \in M} p(n,m).\log_2 p(m|n)$$

where the sum is over the 4 nouns in the set of plural nouns *N* and the 3 markers in the set of possible markers *M* (null, *fip*, *tay*). A language which always uses the same marker for each noun will yield *H*=0, *H*>0 for less predictable languages.

Figure 2 plots conditional entropy against participant number. A repeated measures ANOVA reveals a significant effect of participant number on conditional entropy ($F_{(2.065, 18.464)} = 27.472$, $p < 0.001$, Huynh-Feldt correction). We can also use Page's trend test (Page, 1963) to test the more specific hypothesis that conditional entropy decreases cumulatively across generations (H of initial language > H of participant 1 language > … > H of participant 5 language): this hypothesis is confirmed (L=870.5, m=10, n=6, p<0.001). A typical fifth-participant language exhibits the type of predictable variability described above: for instance, *fip* used to mark plurality on *cow* and *pig*, *tay* used to mark plurality on *rabbit* and *giraffe*.

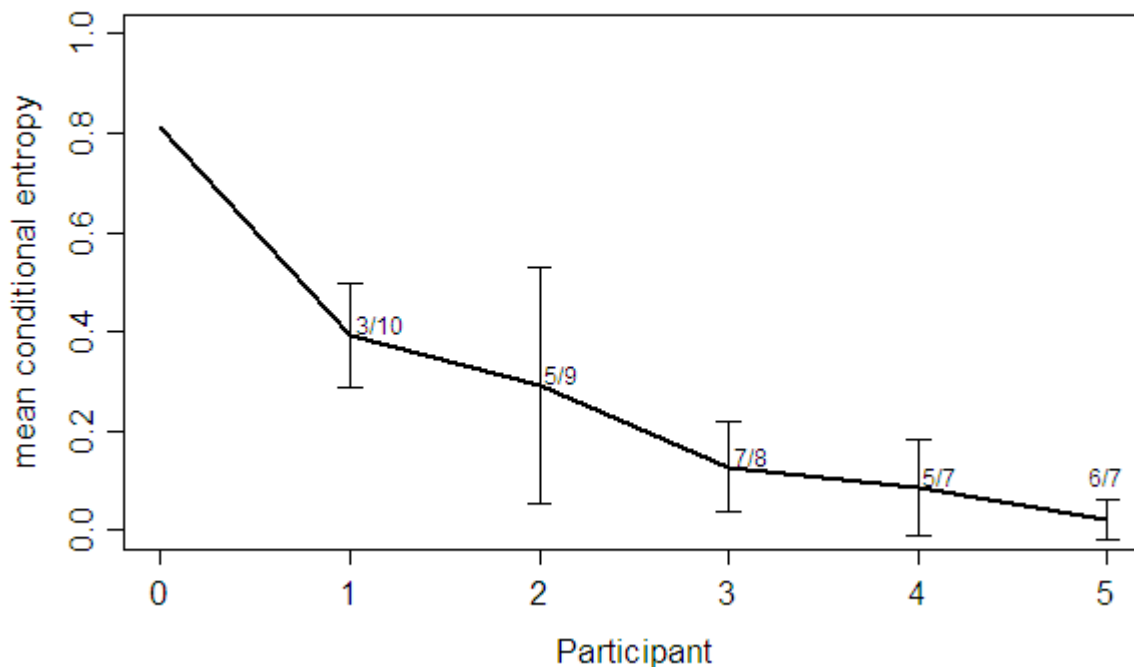**Figure 2:** Conditional entropy of the language produced by each participant (Participant 0 is the input language), averaged over all 10 chains. Error bars give 95% confidence intervals on the mean. Annotations give the number of languages which have significantly non-random use of marking (see text for details) as a proportion of those languages which still use multiple markers for the plural.

While the reduction in conditional entropy effected by the chain-initial participants is significant (mean difference from input language = -0.418, SD=0.17, t(9)=7.795, p<0.001)[7], this is somewhat unsurprising: the initial language exhibits maximal entropy, and participants are likely to reduce entropy if they deviate at all from this language. We can use Monte Carlo techniques to establish whether a given level of entropy associated with a particular distribution of plural marking is likely to arise by a chance assignment of markers to nouns, or whether that level of entropy represents non-random alignment of nouns and markers (i.e. regularization). For each participant's output language, we generated 100,000 random languages which used the same proportion of the various markers but assigned those markers to the plural nouns at random. We measured the conditional entropy of those random languages and compared the resulting distribution to the conditional entropy of the actual language produced by the participant: a participant's language was classified as significantly non-random if it had lower entropy than 95% of the random languages, yielding a one-tailed test with a threshold p=0.05. We conducted this test for each participant whose output language used more than one marker (all one-marker languages have equal [H=0] conditional entropy, rendering this statistic uninformative). The resulting numbers of significantly non-random systems for each language in the diffusion chain are given as annotations on Figure 2: whereas only three of the first participants' languages are significantly non-random,[8] this rises to six of the seven chain-final languages still exhibiting variation in plural marking. This reinforces the claim, consistent with the outcome of Page's trend test, that the elimination of unpredictable variation is cumulative, rather than purely a consequence of the behaviour of the first learner in each chain.

4. Discussion

Simple diffusion-chain populations of adult learners maintain variability in plural marking over repeated episodes of learning, but cumulatively increase the predictability of that variation: the end state in nine of our ten diffusion chains is a language which exhibits no unpredictability, despite six of those nine languages using more than one form to mark the plural. Chains of adults eliminate unpredictability by lexicalising the choice of plural marker: over time, each noun comes to be associated with a particular plural marker.

Previous artificial language learning research has shown that adult and child learners are sensitive to the extent to which input variability is lexically conditioned (Wonnacott, Newport & Tanenhaus 2008; Wonnacott & Perfors, 2009). This information affects the tendency to generalize variants to new lexical items, in line with the predictions of Hierarchical Bayesian models which evaluate word-specific patterns in accordance with higher-level information about variability (Kemp, Perfors & Tenenbaum, 2007; Perfors, Tenenbaum & Wonnacott, in press). The current work suggests that adult learners also have some bias in favour of a lexicalized system. This bias is sufficiently weak that it may not be apparent in isolate learners, but is amplified by the process of iterated learning.

---

[7] An analysis combining the 10 chain-initial participants with the 15 isolates also reveals a statistically-significant drop in conditional entropy, mirroring the results for the chain-initial participants alone: mean difference = -0.399, SD=0.208, t(24)=-9.602, p<0.001.

[8] Three of the 15 isolates produced a language which was significantly non-random according to the Monte Carlo statistic.

It is worth comparing our results with those from Reali & Griffiths' (2009) study of the iterated learning of object-word mappings. In both cases, unpredictability is gradually eliminated. However, the nature of the final system is rather different: whereas Reali & Griffiths observe convergence to a system which does not exhibit variability (one of two possible labels for each object is eliminated), we see stable variability in plural marking. The difference presumably lies in the availability of *context* upon which variability can be conditioned. In Reali & Griffith's study, objects and their labels are presented in isolation: given a learner preference for predictability, there is therefore no way of organising the system such that variability can be preserved. In contrast, in our study the (minimal) linguistic context provided by the noun is sufficient to allow the persistence of conditioned variability. Of course in the real world case multiple conditioning environments (both linguistic and non-linguistic) are available.

These results suggest that cultural transmission in simple diffusion-chain populations may lead to regularization and elimination of unpredictability in languages, even where isolated learners do not exhibit this effect. This result therefore weakens the potential link between strong (child) learner biases against unpredictable variation and elimination of such variation during language change and creolization: given that change and creolization are population-level processes, it may be that weaker (adult) biases in favour of predictability can yield the observed effects. It is important to note that this does *not* imply that regularization is never a one-step consequence of a single learner's behaviour – there is clear evidence that this can happen (e.g. Singleton & Newport, 2004). It is conceivable that certain situations (e.g. involving different types of variation or learners of different ages: Hudson Kam & Newport, 2009) might lead to single step eradication of variation whereas others might elicit more gradual, cumulative elimination in populations.

The model of cultural transmission which we adopt in this work is highly simplified in several respects. Transmission in our diffusion chains is unidirectional, with no reciprocal interaction between individuals. While this simplification allows us to explore the consequences of the minimal level of interaction necessary to support cultural transmission, real language transmission features far richer forms of interaction and there is some evidence that such interaction may be necessary for certain types of communication system to emerge (see e.g. Garrod et al., 2007; 2010). The effect of reciprocal interaction on the regularization of unpredictable variation is currently unknown, although one possibility is that regularization might actually occur more quickly due to alignment during interaction (Branigan, Pickering & Cleland, 2000).

A second important simplification is that, in our diffusion chains, each "generation" consists of only a single individual. While a range of population treatments exist in the diffusion chain literature (see Mesoudi & Whiten, 2008; Whiten & Mesoudi, 2008), there has been little systematic manipulation of population size, and the modelling literature presents a somewhat mixed picture regarding the extent to which results from simple diffusion chains generalise to populations where learners learn from multiple individuals (see Smith, 2009; Burkett & Griffiths, 2010). We are currently exploring the impact of these more complex population dynamics on the process of regularization: however, the work presented here

demonstrates that even very simple treatments of cultural transmission may lead to linguistic outcomes which differ markedly from those observed in isolate learners.

In conclusion, since natural languages are population-level phenomena, population-level processes must be taken into account when considering the creation and maintenance of linguistic predictability. Our approach coincides with a recent series of computational and experimental studies which suggest that the relationship between the prior biases of learners and outcomes of social learning in populations of such learners are non-trivial (Kirby *et al.*, 2007; Kirby *et al.*, 2008; Smith, 2009). Cultural transmission may act to amplify weak biases, and therefore obscure the relationship between the biases of learners and population-level consequences of those biases. The practical consequence of this is that we cannot simply read off the biases of learners from population-level behaviour, nor extrapolate with confidence from individual-based experiments to population-level phenomena: strong constraints at the population level may arise from weak biases which are hard to detect at an individual level.

References

Austin, A. C., Newport, E. L., & Wonnacott, E. (2006). Predictable versus unpredictable variation: Regularization in adult and child learners. Paper presented at the Boston University Conference on Child Language Development, November.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition, 75,* B13-25.

Burkett, D., & Griffiths, T. L. Iterated learning of multiple languages from multiple teachers. In A. D. M. Smith, M. Schouwstra, B. de Boer & K. Smith (Eds.), *The Evolution of Language: Proceedings of the 8th International Conference* (pp. 58-65). Singapore: World Scientific.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31,* 489-509.

Croft, W. (2000). *Explaining Language Change: an evolutionary approach*. London: Longman.

Dixon, R. M. W. (1972). *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

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

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

Garrod, S., Fay, N., Rogers, S., Walker, B., & Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Interaction Studies, 11,* 52-69.

Givón, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition* (Vol. 2) (pp. 1005-1028). Hillsdale, NJ: Lawrence Erlbaum.

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

Griffiths, T. L., Christian, B. R., & Kalish, M. L. (2008). Using category structures to test iterated learning as a method for identifying inductive biases. *Cognitive Science, 32*, 68–107.

Hock, H. H. (2003). Analogical change. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 441-460). Oxford: Blackwell.

Hudson Kam, C., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1,* 151–195.

Hudson Kam, C., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology,* doi:10.1016/j.cogpsych.2009.01.001

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review, 14*, 288–294.

Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10,* 307–321.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, USA, 105,* 10681-10686.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences, USA, 104,* 5241–5245.

Labov, W. (1963) The social motivation of a sound change. *Word, 19,* 273-309.

Lass, R. (1984). *Phonology: An introduction to basic concepts*. Cambridge: Cambridge University Press.

Mather, P.-A. (2006). Second language acquisition and creolization: Same (i-) processes, different (e-) results. *Journal of Pidgin and Creole Languages, 21*, 231-274.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London B, 363,* 3489-3501.

Meyerhoff, M. (2008). Bequia sweet / Bequia is sweet: Syntactic variation in a lesser-known variety of Caribbean English. *English Today, 93,* 31-37.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association, 58,* 216–230.

Perfors, A., Tenenbaum, J.B., Wonnacott, E. (in press) Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*.

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

Siegel, J. (2004). Morphological simplicity in pidgins and creoles. *Journal of Pidgin and Creole Languages, 19*, 139-162.

Singleton, J. L., & Newport, E.L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*, 370-407.

Smith, K. (2009). Iterated learning in populations of Bayesian agents. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 697-702). Austin, TX: Cognitive Science Society.

Whiten, A., & Mesoudi, A. (2008). Establishing an experimental science of culture: animal social diffusion experiments. *Philosophical Transactions of the Royal Society of London B, 363,* 3477-3488.

Wonnacott, E., & Newport, E.L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In A. Brugos, M.R. Clark-Cotton, and S. Ha (eds.), *BUCLD 29: Proceedings of the 29th Annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*, 165-209.

Wonnacott, E. & Perfors, A. (2009) Constraining Generalisation in Artificial Language Learning: Children are rational too. Poster presented at 22nd CUNY conference on human sentence processing.