Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language

CrossMark

Elizabeth Wonnacott [a,*], Helen Brown [b], Kate Nation [c]

[a] Division of Psychology and Language Sciences, University College London, UK
[b] Department of Psychology, University of Warwick, UK
[c] Department of Experimental Psychology, University of Oxford, UK

ABSTRACT

Successful language acquisition requires both generalization and lexically based learning. Previous research suggests that this is achieved, at least in part, by tracking distributional statistics at and above the level of lexical items. We explored this learning using a semi-artificial language learning paradigm with 6-year-olds and adults, looking at learning of co-occurrence relationships between (meaningless) particles and English nouns. Both age groups showed stronger lexical learning (and less generalization) given "skewed" languages where a majority particle co-occurred with most nouns. In addition, adults, but not children, were affected by overall lexicality, showing weaker lexical learning (more generalization) when some input nouns were seen to alternate (i.e. occur with both particles). The results suggest that restricting generalization is affected by distributional statistics above the level of words/bigrams. Findings are discussed within the framework offered by models capturing generalization as rational inference, namely hierarchical-Bayesian and simplicity-based models.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

A classic problem for theories of language acquisition is how learners avoid overgeneralization in the face of an ability to generalize. An example is our knowledge of restrictions on novel combinations of verbs and argument structures, as in the use of "*carry*" in the double-object dative e.g., *"Carry me that"*. Children go through a stage of producing overgeneralizations, yet eventually learn that certain combinations of verbs and structures are restricted. This "paradox" (Baker, 1979) has received a good deal of attention in the literature. Broadly, two different classes of solution have been proposed, one emphasizing increasing knowledge of the semantics of words and constructions (e.g., Pinker, 1989) which eventually provide constraints that block overgeneralizations, and one emphasizing the use of distributional statistics to make inferences about which generalizations are permissible (e.g., Braine, 1971). There is a growing body of evidence suggesting that generalization is constrained by both types of information and that grammatical learning can be characterized as graded rather than absolute (e.g., Ambridge, Pine, Rowland, Freudenthal, & Chang, 2014). This is consistent with the notion that children acquire probabilistic

constraints from input distributions (e.g., Hsu & Chater, 2010; Matthews & Bannard, 2010; Perfors & Wonnacott, 2011; Wonnacott, 2011; Wonnacott, Boyd, Thomson, & Goldberg, 2012).

To inform theory, it is important for experimental work to identify what types of distributional information influence learning and generalization and under which conditions. Here we assess children and adults' sensitivity to a particular distributional property which we term skew. Specifically, we ask whether it is easier to learn arbitrary, lexically based restrictions when structures are not evenly distributed across lexical items (i.e., more words occur with one structure than the other). We also probe the finding from earlier work (Wonnacott, 2011; Wonnacott, Newport, & Tanenhaus, 2008) that it is easier to learn lexical restrictions given broader experience of lexical restrictions within the language.

Artificial language learning provides an ideal tool for exploring learners' sensitivity to different input statistics in isolation of other cues (e.g., semantic, phonological). Wonnacott et al. (2008) took this approach in a series of experiments with adult learners. The input languages incorporated two competing transitive structures and were constructed so that some verbs alternated between structures, but others occurred in just one structure (an arbitrary restriction, since the constructions were synonymous and there were no semantic or phonological cues to verb distribution). Different input sets were used in different learning conditions such that the distributional relationship between verbs and structures was

---

* Corresponding author at: Division of Psychology and Language Sciences UCL, University College London, 2 Wakefield Street, London WC1N 1PF, UK.
E-mail address: e.wonnacott@ucl.ac.uk (E. Wonnacott).

manipulated. Participants were given production and judgment tests after exposure to one of these input sets, and generalization was deemed to have occurred when they produced, or accepted as grammatical, an unattested verb-structure combination. Generalization was found to be affected by the distributional statistics of the learner's input. One factor was verb frequency: verbs frequently encountered in one structure were less likely to be generalized to the other. Importantly, however, participants' learning of verb-structure pairings was affected not only by the frequency of those pairings but also by their more general experience of the language being learned. The likelihood of generalization was influenced by the learners' broader experience of alternation across the input: verbs which had only occurred in one construction were more likely to be generalized to the alternate construction if the learner had experienced more alternating verbs in the input.

Wonnacott (2011) used an adapted learning paradigm to replicate aspects of these findings with 6-year-old children. There are relatively few artificial language learning experiments with children beyond infancy (e.g., Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Hudson Kam & Newport, 2005, 2009; Wonnacott, 2011; Wonnacott et al., 2012). Those that have been conducted indicate that children's learning is substantially slower than that of adults. For example, Wonnacott et al. (2012) found that after three days of training on a single novel verb-argument construction, children produced the structure with correct linking of word order to thematic roles on only 57% of trials, while adults were at ceiling.

The observation that children's learning is slower than adults has implications for experimental design. Unfortunately however, it is not straightforward to simply add additional exposures to compensate for the slower rate of learning. Children can only tolerate short experimental sessions, and schools cannot generally accommodate additional sessions to mitigate this. It is thus necessary to design artificial languages where the "baseline" structures can be acquired relatively quickly. Given these constraints, in order to be able to directly focus on the balance between generalization and lexically specific constraints given relatively little exposure, Wonnacott (2011) used a learning paradigm where the critical relationships were between nouns and meaningless words referred to as "particles", rather than verbs and verb constructions. To facilitate learning, the languages used novel particles but familiar English nouns. This simpler paradigm allowed the same types of statistical manipulations as in Wonnacott et al. (2008) to be explored, with languages containing both alternating nouns (i.e., nouns which occurred with both particles) and nouns restricted to occur with just one particle. A production test was used to probe generalization.

In line with the previous effects of verb frequency in adults, noun frequency played a role, with more generalization to the non-occurring particle for low frequency nouns. Again, however, generalization was also affected by learners' more general experience of the language being learned. Most relevant to the current work, Experiment 1 compared the learning of *minimal-exposure items* in different language contexts. Each of the two *minimal-exposure items* occurred only with one of the two particles, and both were low frequency (presented four times each). The question was whether learners would restrict their usage of *minimal-exposure items* to the particle with which it had occurred in the four exposures, or generalize and extend it to the other particle. From the perspective of individual lexical frequency, four exposures is a very small sample and learners might therefore be expected to ignore this item-specific input and generalize. Importantly, however, these items were introduced later in the experiment, *after* the participants had been exposed to language input containing other nouns. How these *minimal-exposure items* were treated depended critically upon the input to which the children had been previously exposed. Those previously exposed to an input language where each noun occurred with just one of the two particles

(dubbed the *lexicalist* language) were more likely to avoid generalizing with the *minimal-exposure nouns*, treating them as restricted to occur with the one particle with which they were attested. In contrast, learners who had been exposed to a language where all verbs alternated (dubbed the *generalist* language) treated *minimal-exposure nouns* as alternating. Thus children's learning of the restrictions on particular nouns appeared to be affected by their more general learning of how nouns tended to behave across the whole language.

An additional factor explored in the same experiment, and using the same input languages, was whether children could pick up on the statistical prevalence of the particles in the language overall. To this end, in both languages there was a 3:1 bias for one particle, achieved in the *lexicalist* language by having three nouns occur with particle1, and one noun with particle2, and in the *generalist* language by having the 4 alternating nouns each have a bias to occur three times more often with particle1. Testing with entirely novel nouns revealed that children exposed to both *lexicalist* and *generalist* languages had learned the particle1 bias – i.e., they were more likely to generalize that particle. In addition, children in the *generalist* condition were more likely to overgeneralize particle1 with the *minimal-exposure nouns*.

Perfors, Tenenbaum and Wonnacott (2010; Wonnacott & Perfors, 2009) demonstrated that this pattern of learning is in line with the predictions of a hierarchical Bayesian model. This domain general model was originally developed by Kemp, Perfors, and Tenenbaum (2007), who applied it to a set of cognitive learning problems (e.g., acquisition of the "shape bias" in word learning). It is characterized by an ability to track statistical distributions at multiple levels of abstraction (in our work, the distribution of particles used with particular nouns and the language-wide distribution of particles), and to make inferences about the extent to which these levels provide a good indicator of future behavior. This is achieved via the formation of "overhypotheses" about a particular dataset. For example, when trained on the *lexicalist* language, the model formed an "overhypothesis" to the effect that the usage of particles was highly consistent for particular nouns, whereas when trained on the *generalist* language it formed the "overhypothesis" that noun identity and particle usage were unrelated. These "overhypotheses" led to the model showing the same difference in the learning of *minimal-exposure items* as human learners, i.e., greater learning of the associations between these items and their attested particles in the *lexicalist* than *generalist* language. The model also mimicked human performance in showing greater generalization with the more frequent of the two particles/structures, both with novel items and with the *minimal-exposure items* in the *generalist* language. This is due to the fact that it tracked their distribution across the whole language.

The current work builds on previous work by focusing on a property of the lexicalist-input sets used by Wonnacott (2011): the *skewed* distribution of particles across *input nouns*. This skew was originally included to explore the learning of language-wide patterns of particle usage. Potentially however, skew might in itself be an aid to lexical learning. Skewed distributions are common in natural languages. For example, constructions tend to occur more frequently with a single verb (e.g. the double-object [DO]-dative construction occurs more with "give" – "*he gave her the present*" - than with any other construction, and this distribution may benefit learning of its meaning; Casenhiser & Goldberg, 2005; Goldberg, Casenhiser, & Sethuraman, 2004). Another type of skewed distribution is common in grammatical systems where there are alternative forms serving the same function. In this situation, it is often the case that there is one particular form which is used with the majority of lexical items (e.g. the regular English plural –s) while other forms are used with a minority of lexical items (e.g. English plural exception forms such as *feet* and *children*).

The effect of this latter type of skew on the learning of lexical patterns has not been investigated. Intuitively however, it might be easier to learn that particular words are associated with particular structures when there is a majority structure which can act as a "default": once the default has been acquired, associations need only be learned for exception items, whereas if there is no default (for example, if there are two structures which are used equally often) separate associations must be learned for each lexical item. If so, returning to the Wonnacott (2011) languages, it should be harder to learn the association between nouns and particles in a version of the *lexicalist* language without skew. If exposure to *skewed* input aids more general learning about the lexical nature of the language, then we would also expect better learning and less overgeneralization with *minimal-exposure nouns* after *skewed* than *unskewed* input. Moreover, although the *generalist* and *lexicalist* conditions used by Wonnacott (2011) both exhibited equivalent skew of particles across the language, meaning that the greater learning with minimal-exposure verbs could not be due to overall skew per se, the presence of skew in the *lexicalist* language could potentially be a necessary condition to drive learning of the higher order generalization about the lexical nature of the language. To explore this, it is necessary to compare the learning of lexical constraints in an entirely lexical but *unskewed* language, compared to one with alternation. Experiment 1 examined the role of skew in an artificial language learning experiment with 6 year olds and we asked whether adults showed similar patterns of learning in Experiment 2.

Experiment 1

We addressed two central questions. First, does skew aid learning of lexical restrictions? To explore this, two groups of 6-year olds were exposed to languages based on those constructed by Wonnacott (2011). Specifically: (i) *lexicalist-skewed* language, comprising five particle1-only nouns and one particle2-only noun; (ii) *lexicalist-unskewed* language, comprising three particle1-only nouns and three particle2-only nouns. These languages were both fully lexical (no alternating nouns) but if children benefit from skewed input they should show better learning of noun-particle associations in the lexicalist-skewed language (i.e. better learning where there is a majority default particle, used for most nouns, along with an exception form, than when each particle occurs with an equal number of nouns).

We tested learning using two different test types. First, in the *input nouns* test we asked children to produce their own sentences with the trained nouns, i.e. with the nouns which had occurred in the exposure set. If skew plays a role, we predict stronger learning of the restrictions on these nouns (i.e. greater usage of the correct particle rather than the incorrect particle) for children learning the *lexicalist-skewed* language compared with the *lexicalist-unskewed* language. Second, following Wonnacott (2011), we also asked children to produce sentences with two "minimal-exposure" nouns that were introduced in an additional exposure session occurring only *after* exposure to the main language input. In the additional exposure session, which was identical across conditions, each minimal-exposure noun occurred just four times, with one of the minimal-exposure nouns always occurring with particle1, and the other always occurring with particle2. Children were then asked to use these two nouns in their own sentences, and we looked to see whether they continued to use them with the attested particle. This test allowed us to see whether exposure to skewed input was sufficiently general to aid ongoing learning of lexical restrictions. We predicted stronger learning of the restrictions on the minimal-exposure nouns (i.e. continuing to use them with the particle with which they occurred in the four exposures, rather than generalizing to the unattested particle) for participants

previously exposed to the *lexicalist-skewed* input than for those who were exposed to the *lexicalist-unskewed* input. Note that minimal exposure items are more appropriate than entirely novel nouns since they allow us to look at how learners balance generalization against some minimal lexically based information.

Our second question was whether there is a benefit of overall lexicality, even in the absence of skew. To explore this, two further groups of 6 year olds were trained and tested on languages to be compared with the *lexicalist-unskewed* language, specifically: (i) *mixed* language, comprising one particle1-only noun, one particle2-only noun and four (unbiased) alternating nouns and (ii) *generalist-unskewed* language, comprising six (unbiased) alternating nouns. Neither of these languages contained skew, but if overall lexicality aids learning, children should find it easier to learn restrictions on nouns in the *lexicalist-unskewed* language, than in either of these two languages where nouns can alternate. Again, learning was probed with two types of test items. First, in the *input nouns* test, children produced sentences with nouns from the exposure set. Note that here, since only non-alternating nouns are relevant (because we are specifically interested in learning of the restrictions), and since there are no non-alternating nouns in the *generalist-unskewed* condition, only the *lexicalist-unskewed* language and *mixed* language are compared, with greater learning predicted in the *lexicalist-unskewed* language. Second, there was the *minimal-exposure nouns* test, where participants produced sentences with the two nouns which had been presented just four times (as *particle1-only* and *particle2-only*) in an additional exposure session. If lexicality plays a role, we predict stronger learning of the restrictions on the new *minimal-exposure* nouns for learners who were previously exposed to the *lexicalist-unskewed* language than those who were exposed to either the *generalist-unskewed* or *mixed* languages. Full details of the languages and test items are described in Table 2.

We also determined whether children had any explicit awareness of their learning of lexicality via a post-experiment interview. The relationship between implicit and explicit learning in artificial language learning experiments within the statistical learning literature is not well understood, although there seems to be an assumption that it is largely implicit, at least in children. Collecting subjective data is a first step towards exploring this issue.

Method

Participants

Data were collected from sixty 6-year-old children. Participants were monolingual native English speakers with no known hearing, language, or speech disorders. Two of the original 60 children were replaced as they failed to contribute any data which met baseline performance (see Results). Each child was randomly assigned to one of four conditions (see Table 1). Since our contrasts compare the *lexicalist-unskewed* condition against each of the other conditions, we used t-tests to compare the mean age and listening span of this group against those of each of the other groups - no significant differences were found (all $ps > .2$).[1] Informed consent was

---

[1] Although there were no significant differences between conditions, an anonymous reviewer pointed out that the condition where we see strongest learning (the skewed condition), is also the one in which the children are oldest and have the largest listening span. We therefore conducted a series of additional analyses to explore this confound. These analyses are included at the end of the R script provided online at http://rpubs.com/ewonnacott/235483. In sum, there were no reliable correlations with either listening span or age in our data, and adding these factors into the linear mixed effects models did not change the pattern of results (the difference between the lexicalist skewed and lexicalist unskewed conditions remained significant). In addition, removing the three oldest children from the lexicalist skewed condition (so that age was matched across conditions) also did not change the pattern of results.

Table 1
Mean (*SD*) age and listening span standard score of children in Experiment 1 as a function of input condition. Standard score population *M* = 100, *SD* = 15.

Input condition	*N* total	*N* male	Age	Listening span[a]	Counterbalancing
Lexicalist-skewed	15	12	6;3 (0;4)	109 (10)	8 maj. = *dow*
					7 maj. = *tay*
Lexicalist-unskewed	15	9	6;2 (0;5)	107 (20)	NA
Mixed	15	8	6;0 (0;6)	102 (15)	NA
Generalist-unskewed	15	7	6;2 (0;4)	107 (21)	NA

[a] Automated Working Memory Assessment (Alloway, 2008).

Table 2
Noun types included in input and tests.

	Training/Test	Lexicalist-skewed		Lexicalist-unskewed	Mixed	Generalist-unskewed
		version1 (*n* = 7)	version2 (*n* = 8)			
Input noun 1	Input training and input test	tay-only	dow-only	dow-only	dow-only	Alternating
Input noun 2	Input training and input test	dow-only	tay-only	tay-only	tay-only	Alternating
Input noun 3	Input training and input test	dow-only	tay-only	dow-only	Alternating	Alternating
Input noun 4	Input training and input test	dow-only	tay-only	tay-only	Alternating	Alternating
Input noun 5	Input training	dow-only	tay-only	dow-only	Alternating	Alternating
Input noun 6	Input training	dow-only	tay-only	tay-only	Alternating	Alternating
Minimal-exp. noun1	Minimal exp. training and minimal exp. test	dow-only	dow-only	dow-only	dow-only	dow-only
Minimal-exp. noun2	Minimal exp. training and minimal exp. Test	tay-only	tay-only	tay-only	tay-only	tay-only

obtained from both schools and parents prior to the start of the experiment.

Stimuli

Stimuli were sentences that began with the word *moop*, followed by one of 16 English nouns with familiar referents (*bee, camel, donkey, duck, frog, giraffe, hippo, kangaroo, monkey, owl, parrot, penguin, pig, rabbit, tiger, zebra*) and one of two sentence-final novel particles (*dow, tay*). Sentences took the form "*moop noun dow/tay*", where *moop* was intended to mean "there are two" (following Wonnacott (2011), and chosen since plurality is a salient property and simple to depict). Stimuli were recorded by a female native British English speaker. Words were edited into separate sound files and peak amplitude was normalized using Audacity (http://audacity.sourceforge.net/). Clipart pictures of the 16 noun referents were selected from free online databases and were edited to generate pictures that showed pairs of items (e.g., two tigers, two penguins, *etc.*).

The structure of the four input languages is summarized in Table 2. Four of the six *input nouns* (labeled as input nouns 1–4 in Table 2) featured in the *input nouns* test, a production test which immediately followed input training (note: two *input nouns* were not tested – this was to provide children with the opportunity to produce multiple sentences with the same nouns yet avoid overlengthy testing). This test was identical across conditions, although the children's experience with the nouns, and whether they had been restricted or alternating, differed across conditions. Only data from nouns that had been particle1-only and particle2-only were analyzed, meaning that there were different numbers of test items across conditions, and no data at all in the *generalist-unskewed* condition (the test was included for consistency across conditions). However a second test was also included where data from all conditions were analyzed. This was the *minimal-exposure nouns* test. This test featured two new nouns which were presented to the learner in a short exposure session (minimal-exposure training) administered immediately before the test. Importantly, this exposure was identical across the conditions and comprised one noun that occurred four times with particle1, and a second that occurred four times with particle2. Thus the *minimal-exposure nouns* test explored whether learning of the restrictions on new low fre-

quency particle1-only/particle2-only nouns differed for learners previously exposed to different input languages.

The six *input* and two *minimal-exposure nouns* were randomly selected for each participant from the set of 16 possible nouns to avoid the possibility of item-based effects. Assignment to *dow* versus *tay* as the minority or majority particle was counterbalanced in the *lexicalist-skewed* condition.

Procedure

Children were tested individually in a quiet area of their school. Tasks were run on a Toshiba laptop using ExBuilder software, a custom built software package developed at the University of Rochester. Each child completed three experimental sessions, the majority on three consecutive days though 11 children (3 × *lexicalist-skewed*, 4 × *lexicalist-unskewed*, 3 × *mixed*, 1 × *generalist-unskewed*) completed the three sessions over four or five days due to absence from school on one or more days.

Fig. 1 summarizes the tasks and testing schedule. Children were introduced to a toy elephant at the beginning of Session 1, and told were told that they were going to learn how to say some things like "Ellie Elephant".

Noun practice. In Session 1 children completed two noun practice tasks. First, they viewed a picture (e.g., one tiger) while hearing its English name ("*tiger*") and repeated the name aloud. Second, they viewed the same pictures and were asked to produce the corresponding names on their own. This second task was repeated at the beginning of Sessions 2 and 3 to ensure that the children labeled the pictures correctly. When incorrect labels were provided the participant was told "*This one is called a tiger. Can you say tiger?*". Trial order was randomized on a child-by-child basis.

Input nouns training. Children heard 12 sentences per block of training, with each of the six nouns being heard twice per block. On each trial the child saw a picture (e.g., two tigers), heard a sentence (e.g., *moop tiger dow*) and repeated the sentence aloud. If any element of the sentence was mispronounced the experimenter said "*Almost, this one was 'moop tiger dow'. Can you say that?*". If a sentence was mispronounced a second time no feedback was provided and the next trial initiated. Trial order was randomized on a child-
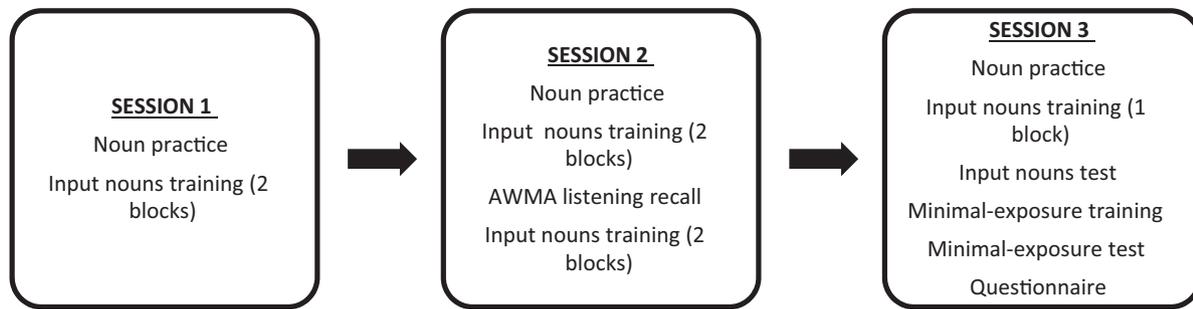
Fig. 1. Summary of tasks completed in each of the three experimental sessions.

by-child basis such that the same animal was not presented twice in a row.

Input nouns test. Children saw a picture (e.g., two tigers), heard the first word of the sentence ("*moop*"), and were asked to complete the rest of the sentence on their own. If the noun was produced incorrectly they were given corrective feedback ("*Good try, but this one is a tiger, not a lion*") and asked to say the sentence again using the correct noun. These trials were not included in the analyses. No corrective feedback was provided regarding the usage of sentence-final particles. In order to keep the test of reasonable length, only nouns 1–4 from the input training task (see Table 2) were encountered at test with each presented four times. Trial order was randomized on a child-by-child basis such that the same animal was not presented twice in a row.

Minimal-exposure nouns training. Two further nouns that had not featured in the input nouns training were encountered during the minimal-exposure nouns training. One always occurred with *dow* and one with *tay*. As in the input nouns training task, children saw a picture of the noun, heard "*moop noun dow/tay*", and repeated the sentence aloud. Each *minimal-exposure noun* was encountered four times with feedback provided where necessary, as per the input nouns training task. Trial order was randomized for each child with no constraints (since there were only two animals a fully random order was preferable to one in which the two sentences alternated).

Minimal-exposure nouns test. This was identical to the *input nouns test*, with pictures of the two *minimal-exposure nouns* being presented and children required to complete the sentence (given the initial word "*moop*"). Each noun was tested during the first or second trial, then three repetitions of the two nouns occurred in a random order. There were four trials per noun.

Questionnaire. At the end of the final session the experimenter interviewed each child to ascertain any patterns that they had noticed in the experimental language. We asked them to describe how they knew when to use dow/tay, and if they noticed any patterns in the way Ellie Elephant used them. Of interest was whether children showed any awareness of the fact that particle usage could be conditioned on the noun (e.g., "*donkey goes with dow*").

Results

Results from the *input nouns* and *minimal-exposure nouns* tests were analyzed separately. We were interested in looking at the learning of the restrictions on non-alternating nouns, therefore for the *input nouns test*, we only analyzed data for nouns which were restricted in the input, that is, those that were dow-only or tay-only (meaning that for this test, no data were analyzed for

the *generalist-unskewed* condition). Trials were excluded if children had initially used an incorrect noun and been corrected by the experimenter, if they inserted an alternative word for a particle, or if they failed to include a particle. Children were not penalized for omitting to repeat the initial word *moop*. The proportion of trials failing to meet these baseline criteria is reported below. Data were analyzed using both frequentist and Bayesian methods. (Note that further information on excluded data can be viewed online at http://rpubs.com/ewonnacott/242454; this script includes information about error trials (the frequency of "other particle" and "no particle" trials) and presents the patterns of particle usage for the *alternating* nouns along with some basic analyses in terms of regularization – cf. Hudson Kam & Newport, 2005; Samara, Smith, Brown, & Wonnacott, in press).

For frequentist analyses, since the dependent variable was binary (i.e. whether the particle in the response was correct/incorrect) the data were analyzed using logistic mixed effects models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008; Quené & Van den Bergh, 2008) in the package lme4 (Bates, Maechler, Bolker, & Walker, 2013) for the R computing environment (R Core Team, 2012). These models allow binary data to be analyzed with logistic models rather than proportions, as recommended by Jaeger (2008). For each of our analyses, the dependent variable was whether the correct (i.e. attested) particle was produced (coded as 1 vs. 0). The independent variable was *condition* which had either three levels (input nouns analyses: *lexicalist-skewed*, *lexicalist-unskewed* and *mixed*) or four levels (minimal-exposure nouns analyses: *lexicalist-skewed*, *lexicalist-unskewed mixed* and *generalist-unskewed*). In each case, *lexicalist-unskewed* was the reference level so that we could inspect the contrast between this condition and the other conditions. This was achieved within the lme4 package by replacing the three-way (or four-way) factor "condition" with two (or three) centered dummy variables and using the main fixed effects from the output of this model. We used this technique throughout. We also included whether the correct particle for the trial was *dow* or *tay* as a control variable, as well as the interaction of this with condition. There was no effect of these control factors in any of the models, therefore although they were retained in the model they are not reported. All variables were coded as centered, numerical predictors so that effects in the model could be evaluated as the average effects over levels of the other predictors. To avoid anti-conservative conclusions (Barr, Levy, Scheepers, & Tily, 2013), we specified a full random effects structure in our models, including intercepts for subjects and by-subject random slopes for all within-subject factors and their interactions. All models converged with Bound Optimization by Quadratic Approximation (BOBYQA optimization; Powell, 2009).

For Bayesian Analyses, we computed Bayes factors using the method advocated by Dienes (2008, 2015). This requires an estimate of the predicted mean difference between conditions according to H1. Recall that our key aim is to de-confound the benefit that

Wonnacott (2011) saw for the learning of lexically restricted nouns after exposure to the *lexicalist* language compared to the *generalist* language: was this due to the lexical nature of the input, and/or did skew play a key role? We address this by contrasting an unskewed version of the lexicalist language (*lexicalist-unskewed*) with matched languages where lexicality and skew are manipulated separately. Wherever we contrast our *lexicalist-unskewed* condition with each of the other conditions, we inform our H1 by the difference between the *lexicalist* and *generalist* conditions for minimal-exposure nouns in Wonnacott (2011). In that experiment, participants produced the correct particle for minimal-exposure nouns 86% of the time in the *lexicalist* condition and 66% of the time in the *generalist* condition. However, so as to meet assumptions of normality, we work in log-odds space. To obtain the estimate of predicted difference, we ran a logistic mixed effects model equivalent to those reported in the current paper, over the relevant data from Wonnacott (2011). The estimate obtained was 2.758. Following Dienes (2008) we model H1 by using this estimate as the *SD* of a half normal distribution.

For each of the comparisons between conditions, we used Bayes factors (BF) to test the strength of evidence for this model compared with a null hypothesis of no difference between conditions. Our sample estimate is the estimate produced for the coefficient from the relevant lme model used in the frequentist analyses described above. Following Dienes (2008), a BF of 3 or above is taken to indicate substantial evidence for the alternative rather than the null hypothesis, while a BF of 1/3 or below is taken to indicate substantial evidence for the null rather than alternative hypothesis. Thus, a BF between 3 and 1/3 indicates data insensitivity for distinguishing the alternative and null hypotheses (see Dienes, 2008, 2014). Full details of analyses can be found in the R analyses script which is available online at http://rpubs.com/ewonnacott/242454. Data are also available at https://osf.io/2zfe6/.

Input nouns

Data from the *lexicalist-skewed, lexicalist-unskewed* and *mixed* conditions were analyzed (recall that there were no non-alternating nouns in the *generalist-unskewed* language). There were more contributing data points per child in the *lexicalist-skewed* and *lexicalist-unskewed* conditions than in the *mixed* condition (since in the latter, two of the nouns being tested were alternating nouns); however logistic linear mixed effects models are robust to problems associated with proportional data from uneven samples. Trials not meeting the baseline criteria described above were excluded (*lexicalist-skewed*: 3%, all majority nouns; *lexicalist-unskewed*: 5%; *mixed*: 2%).

The proportion of correct responses in each condition is shown in Table 3. Our first hypothesis was that skew would benefit learning, leading to the prediction that more correct particles would be produced overall in the *lexicalist-skewed* condition than the *lexicalist-unskewed* condition. The data were consistent with this prediction (*M skewed* = 90% and 99% for the minority and majority

particles respectively; *M unskewed* = 74%; beta = 3.11, SE = 0.74, z = 4.20, p < .001; BF = 1974.81). Our second hypothesis was that the overall lexicality of the language should lead to more productions with the correct particle in the *lexicalist-unskewed* than the *mixed* condition (recalling that we look only at the two non-alternating nouns in the latter language). In fact, there was evidence for no difference (*M lexicalist-unskewed* = 74%; *M mixed* = 70%; beta = −0.05, SE = 0.58, z = −0.09 p = .93, BF = 0.22. This suggests that skew but not overall lexicality of the language is helpful to children's learning of the lexical restrictions on these nouns.

Within the *lexicalist-skewed* language, performance on nouns that occurred with the majority particle was higher than performance on the minority particle (90% vs. 99%). Although this pattern of results was predicted, given the sensitivity to overall particle frequency found in Wonnacott (2011), the difference is not significant here, with the BF comparison telling us that the test is insensitive (beta = −1.86, SE = 3.67, z = −0.51 p = .61; BF = 1.07; for this BF comparison we take H1 to be scaled by an estimate of the difference in performance for the majority and minority particle in the *lexicalist* language in Wonnacott (2011), obtained by running an equivalent logistic mixed effects model on that data). However, the children are near ceiling with the majority particle. Note that for the key comparison between the *lexicalist-skewed* and *lexicalist-unskewed* languages, particle frequency cannot lead to overall greater performance in the *skewed* language since although it aids performance with the majority noun it should equally hinder performance with the minority noun. Thus if particle frequency were the only factor at play, performance on the minority particle noun in the *skewed* language should be *lower* than performance on the unbiased nouns in the *unskewed* language, which is not what we see. Nevertheless, a potential concern is that more majority nouns than minority nouns were included. We thus repeated the analysis with just one majority and one minority noun included (achieved by removing *input nouns* 3 and 4 from the *skewed* language condition). Performance in the *skewed* language remained high (95%, SE = 3%) with evidence for a difference between the *skewed* and *unskewed* languages (beta = 3.00, SE = 0.84, z = 3.59, p < .001; BF = 209.11).

Minimal-exposure nouns

Data from all four conditions were analyzed. We again excluded trials not meeting the baseline criteria (*lexicalist-skewed*: 6%; *lexicalist-unskewed:* 5%; *mixed: 5%; generalist-unskewed*: 5%). There was no difference in amount of data excluded across conditions (logistic linear mixed effects model, all *ps* > .5).

We again explored two predictions. First, if skew helps children learn new restrictions, more correct particles should be produced in the *lexicalist-skewed* than *lexicalist-unskewed* condition. This prediction was confirmed (*M skewed* = 77% and 92% for the minority and majority particles respectively; *M unskewed* = 64%; beta = 2.66, SE = 0.93, z = 2.88 p = .004; BF = 26.13) demonstrating that, as for the input nouns, skew helps children to learn lexical restrictions.

Table 3

Mean (*SE*) proportion of trials where the attested particle was produced as a function of language input condition for input nouns and minimal-exposure nouns in Experiments 1 and 2.

Input condition	Experiment 1 (Children)		Experiment 2 (Adults)	
	Input nouns	Minimal-exposure nouns	Input nouns	Minimal-exposure nouns
Lexicalist-skewed				
Minority particle	0.90 (0.06)	0.77 (0.11)	0.88 (0.07)	0.93 (0.07)
Majority particle	0.99 (0.01)	0.92 (0.06)	0.94 (0.03)	1.0 (0)
Lexicalist-unskewed	0.74 (0.04)	0.64 (0.08)	0.78 (0.06)	0.89 (0.07)
Mixed	0.70 (0.07)	0.58 (0.07)	0.61 (0.04)	0.84 (0.07)
Generalist-unskewed	N.A.	0.58 (0.07)	N.A.	N.A.

Table 4
Number of children showing some awareness of the fact that particle usage can be conditioned on noun identity.

Input condition	Participants demonstrating some awareness			Participants not demonstrating some awareness		
	N	Mean (SE) input nouns	Mean (SE) min. exposure nouns	N	Mean (SE) input nouns	Mean (SE) min. exposure nouns
Lexicalist-skewed	8	0.94. (0.03)	0.89 (0.07)	7	0.99 (0.01)	0.82 (0.09)
Lexicalist-unskewed	3	0.75 (0.12)	0.67 (0.22)	12	0.73 (0.04)	0.63 (0.09)
Mixed	4	0.63 (0.10)	0.48 (0.02)	11	0.73 (0.08)	0.61 (0.06)
Generalist-unskewed	4	N.A.	0.46 (0.17)	11	N.A.	0.62 (0.07)

Second, if the overall lexicality of the language during training assists learning, children should produce more attested particles in the *lexicalist-unskewed* condition, relative to the *mixed* and *generalist-unskewed* conditions. While the means in Table 3 are consistent with this general pattern, the differences between the conditions were not significant with no evidence one way or the other for a difference between performance following exposure to the *lexicalist-unskewed* language (*M* = 64%) and either the *mixed* (*M* = 58%; beta = −0.49, SE = 0.82, z = −0.60, p = .55; BF = 0.48) or *generalist-unskewed* (*M* = 58%; beta = −0.66, SE = 0.82, z = −0.80 p = .43; BF = 0.60) language.

Again note that, as with the input nouns, accuracy was higher for majority compared to minority particle nouns (92% vs. 77%) though again this test was insensitive (beta = −2.19, SE = 4.72, z = −0.46, p = .64, BF = 1.06; here H1 for the BF analyses was scaled by an estimate of the difference in performance for the majority and minority particle with minimal exposure nouns in the *lexicalist* and *generalist* languages in Wonnacott (2011), obtained by running an equivalent logistic mixed effects model on that data). Note again though that even if there is a benefit of frequency for the majority particle, overall greater performance in the *lexicalist-skewed* language *cannot* be due to particle frequency, since this would have led to an equivalent *decrease* in performance on the minority nouns, which was not seen. Note also that in these analyses there were an equal number of minority and majority particle test items.

Post-experiment questionnaire data assessing explicit awareness

Responses were binary coded to indicate whether the children showed any explicit awareness that particle-usage could be conditioned on the noun. To be coded as aware a child had to mention one (or more) of the noun-particle relationships in their input (e.g., "*donkey goes with dow*").

Table 4 shows the number of children in each condition who showed some awareness, as well as the mean score on the *input nouns* and *minimal-exposure nouns* tests broken down for children who did/did not show awareness in each condition. It can be seen that although the majority of children did not give responses that indicated awareness, more did so in the *lexicalist-skewed* condition than in the other conditions. However chi square tests comparing the *lexicalist-skewed* condition against the other conditions did not provide evidence for reliable differences (ps > .1). Looking at performance within the *lexicalist-skewed* language, there does not appear to be any evidence that participants showing awareness of lexical conditioning outperformed those who did not, with either *input* or *minimal-exposure* nouns, though our numbers are small here.

Discussion

The results from Experiment 1 clearly demonstrate that 6-year-olds benefit from the presence of skew when learning lexically based co-occurrence relations. Children exposed to an artificial language where five nouns occurred with a majority particle and a single noun occurred with a minority particle were more likely to reproduce the correct noun-particle pairings than children

exposed to a language where an equal number of nouns occurred with each particle. This was despite that fact that the frequency of occurrence of the noun-particle bigrams was matched across the two languages. In addition, previous exposure to skewed input conferred an ongoing advantage to the learning of new nouns introduced later in the experiment under conditions of minimal exposure. Skew was not manipulated for minimal exposure nouns, but nevertheless those children previously exposed to *skewed* input were more likely to learn and produce the appropriate particle. Once again, this was not a consequence of bigram frequency which was equivalent across conditions. Rather, children learned differently from this matched exposure depending upon their past experience. Note also that for both *input nouns* and *minimal-exposure nouns*, greater learning in the *skewed* condition cannot be due to simple lexical frequency, i.e., the greater frequency of the majority particle in the *skewed* language. If this was the case, we would have observed better performance only on the nouns taking the majority particle, with equivalent *reduced* performance on the noun taking the minority particle. This is not what we observed.

Contrary to our predictions, we saw no evidence that lexicality of the input had an effect in the absence of skew. That is, children did not show better learning of noun-particle pairings in the *lexicalist-unskewed* language compared with the *mixed* language, despite the presence of alternation in the latter language. For input nouns, our Bayesian analyses suggested that we had substantial evidence for the null hypothesis (i.e. evidence of no difference between conditions for these nouns). For minimal exposure nouns, where we expected previous exposure to alternation to reduce learning and thus have a benefit in the *lexicalist-unskewed* condition, there was also no evidence of a difference between conditions. Here, however, the Bayesian analyses suggested that for these items our test was insensitive. Note that this is *not* the same as finding evidence of "no" difference. We considered whether we had sufficient power to find an effect. If our true mean between conditions was actually zero, based on our current level of variance, N = 22 is required (assuming variance is proportional to the square root of standard error); on the other hand, if we assume that our current estimate was correct, with this level of variance we would need around 17 times as many participants as we do at present to show the effect (261 children per condition). Testing this number is clearly impractical.[2]

At least for the input nouns, children showed reliable evidence of an effect of skew and reliable evidence of *no* benefit of lexicality. This suggests that the benefit of witnessing lexicalist input in Wonnacott (2011) may actually have been dependent on the presence of skew in that input language. These findings have

---

[2] We did have the opportunity to test an additional 11 children in the lexicalist-unskewed and 13 in the mixed conditions. These participants were *not* tested as part of our original planned sample, which was 15 children per condition. Due to practicalities in terms of sample availability at the time of testing, the ages in the new sample were not well matched to those in our original (the children were significantly younger) and we have therefore chosen not to include their data in the reported analyses. We did, however, repeat all of the Bayesian statistics for this larger sample (noting that, in contrast to the interpretation of p-values in Frequentist analyses, Bayes Factors remain a valid measure of evidence even with optional stopping, Dienes, 2016; Rouder, 2014) and the pattern of results was unchanged.

implications with regard to the applicability of the hierarchical Bayesian model – an issue we return to in the General Discussion; we will also consider the implications of the relationship between awareness and condition. First, however, we consider whether the current pattern of results also holds for learning by adults.

Experiment 2

Previous experiments demonstrate that adult learners are sensitive to overall lexicality. Wonnacott et al. (2008, Experiments 1 and 2) devised languages with alternating and one-construction verbs in the input. The two transitive structures were equally balanced, but the proportion of alternating verbs differed across the experiments (4/12 in Experiment 1 vs. 8/12 in Experiment 2). Learning of the constraints on the one-construction verbs was greater in Experiment 1 than Experiment 2. In addition, Wonnacott et al. demonstrated that the learning of the constraints on one-construction minimal-exposure verbs (i.e., verbs exposed four times after the main exposure set) was stronger after exposure to lexicalist input (where all verbs were one construction) than after exposure to generalist input (where construction frequencies matched the lexicalist input, but all verbs alternated). This was demonstrated both under conditions of skew (lexicalist input = 7 verbs occurring in one structure, 1 verb in the other construction; Wonnacott et al., 2008, Experiment 3) and for unskewed input (lexicalist input = 4 verbs in each structure; Wonnacott, Perfors, & Tenenbaum, 2008). Finally, a more recent adult artificial language learning study by Perek and Goldberg (2015) followed up Wonnacott et al. (2008). This experiment again explored the learning of two novel word order constructions and contrasted learning of a fully lexicalist input language (three verbs used consistently with structure 1, three used consistently with structure 2) with one where the frequency of the two constructions was matched but where some verbs (two out of six) alternated. A key difference from the previous work was that these constructions had subtly different functions from each other, a factor which was predicted to encourage generalization, even after exposure to fully lexical input. This prediction was born out, with participants in the lexicalist group displaying some tendency for generalization. However, critically for current purposes, there was nevertheless still *less* generalization for those participants in the lexicalist group than for those who had witnessed alternation. Note that this occurred despite that fact that the constructions were of equal frequency in both languages (i.e. no skew).

Taken together, the findings reviewed above indicate that overall lexicality of the language can aid learning of constraints in adults, even in the absence of skew. This is at odds with the findings from children in Experiment 1, where we saw that, at least for input nouns, children were unaffected by witnessing some alternating nouns in the input (equivalent performance in the lexicalist-unskewed and mixed languages). This might reflect differences between child and adult learners. Alternatively however, it might reflect differences in methodology: the experiments with adults used fully artificial languages (all novel words) where the critical relationships were between verbs and two novel transitive constructions, rather than nouns and particles. To address which explanation is correct, Experiment 2 used the *lexicalist-skewed, lexicalist-unskewed* and *mixed* conditions from Experiment 1 with adults, using identical stimuli and following the same procedures.

Method

Participants

Forty-five undergraduate students (mean age 22 years) from the University of Warwick participated in Experiment 2. Fifteen

were randomly assigned to each of the *lexicalist-skewed, lexicalist-unskewed* and *mixed* conditions. All were monolingual native English speakers with no known hearing, language, or speech disorders. Informed consent was obtained at the start of the experiment.

Stimuli, design and procedure

The materials and procedure were identical to Experiment 1 except that adults completed a written post-experiment questionnaire rather than being interviewed. This included an open-ended question which asked them to describe the structure of the language they had been learning. Responses were binary coded (aware/unaware) according to whether they showed any awareness of the fact that particle usage could be lexically conditioned. This included both mentioning any of the particular animal-particle pairings in their input (or in the *minimal-exposure nouns*) or writing something which indicated awareness that particular animals and particles co-occurred (for example, some participants suggested that particles indicated "gender", presumably on the basis of previous experience with modern foreign languages; this was not seen in the children's data in Experiment 1).

Results

Data analysis procedures and baseline criteria for inclusion were identical to those in Experiment 1. All participants and trials met the baseline criteria and so all were included in the analyses.

Input nouns

The proportion of correct responses in each condition is shown in Table 3. The data were consistent with both of our original hypotheses: there were more correct productions in the *lexicalist-skewed* condition relative to the *lexicalist-unskewed* condition (M skewed = 88% and 94% for the minority and majority particles respectively; M unskewed = 78%; beta = 1.76, SE = 0.69, z = 2.54, p = .01, BF = 10.16), and more correct productions in the *lexicalist-unskewed* condition than in the *mixed* condition (M lexicalist-unskewed = 78%; M mixed = 61%; beta = −1.30, SE = 0.62, z = −2.10, p = .04; B = 3.49). This suggests that adults' learning of lexical restrictions is affected by both skew and overall lexicality of the language. As with the children in Experiment 1, there were more correct productions with the majority particle noun than the minority particle nouns within the skewed language (88% vs. 94%) although the difference was not reliable with the BF comparison telling us that the test is insensitive (beta = 2.43, SE = 3.39, z = 0.72, p = .48; BF = 1.18). To ensure that the difference between the *lexicalist-skewed* and *lexicalist-unskewed* conditions was not carried by the greater number of majority test nouns, we repeated the analyses with input nouns 3 and 4 (see Table 2) removed from the *skewed* language (so that data from just one majority and one minority noun were included), as in Experiment 1. Performance in the *skewed* language remained high (94%, SE = 3%) with evidence for a contrast between the *skewed* and *unskewed* languages (beta = 2.15, SE = 0.81, z = 2.66 p = .008; BF = 14.55).

Minimal-exposure nouns

Once again the pattern of performance (Table 3) was in the predicted direction: *lexicalist-skewed*: minority particle: 100%, majority particle: 93%, *lexicalist-unskewed*: 89%, *mixed*: 84%. A model with the same structure as previous models showed no significant difference for either comparison, with BF's indicating that we did not have sufficient sensitivity to distinguish H1 and H0 in either case (*lexicalist-skewed* versus *lexicalist-unskewed*: beta = 10.92, SE = 41.58, z = 0.26, p = .79; BF = 1.01; *mixed* versus *lexicalist-unskewed*: beta = −0.70, SE = 1.99, z = −0.35 p = .72; BF = 0.75). We discuss possible reasons for this lack of an effect in the Discussion.

Table 5
Number of adults showing some awareness of the fact that particle usage can be conditioned on noun identity.

Input condition	Participants demonstrating some awareness			Participants not demonstrating some awareness		
	N	Mean (SD) input nouns	Mean (SD) min. exposure nouns	N	Mean (SD) input nouns	Mean (SD) min. exposure nouns
Lexicalist-skewed	14	0.92 (0.04)	0.96 (0.04)	1	1 (0)	1 (0)
Lexicalist-unskewed	15	0.78 (0.06)	0.89 (0.07)	0	N.A.	N.A.
Mixed	9	0.61 (0.06)	0.86 (0.06)	6	0.60 (0.08)	0.81 (0.14)

Post-experiment questionnaire data assessing explicit awareness

Table 5 shows the number of adults in each condition who showed some awareness that particle usage was conditioned to lexical items, as well as the mean score on the *input nouns* and *minimal-exposure nouns* at test, broken down for participants who did/did not show awareness. In contrast to children, the majority of adults were coded as aware, although there were more unaware participants in the *mixed* condition. Fisher exact analyses suggested that the difference between conditions was reliable for *mixed* versus *lexicalist-unskewed* ($p = .017$), and marginal for *mixed* versus *lexicalist-skewed* ($p = .081$). While there is very little data, looking specifically within the *mixed* condition, we see little evidence for a relationship between their awareness of lexicality as so measured, and their usage of correct particles.

Discussion

Like the children in Experiment 1, adults showed better learning of noun-particle pairings when exposed to an artificial language in which five nouns occurred with a majority particle and just one occurred with a minority particle, compared to a language where an equal number of nouns occurred with each particle. However, adults, unlike the children in Experiment 1, showed evidence of benefitting from overall lexicality: they showed stronger learning of noun-particle relationships in the *unskewed* but entirely lexically consistent language, relative to a language in which particle usage was consistent for two nouns but alternated for the other four. This is consistent with the pattern of results seen during verb construction learning in *unskewed* languages by adults (Wonnacott et al., 2008), and with the predictions of a hierarchical Bayesian model (Perfors et al., 2010).

Turning to the *minimal-exposure nouns*, there were no reliable differences across conditions. Moreover, the Bayes factor analyses suggested that our test is actually insensitive for both noun types. In retrospect, we think that our paradigm here is not optimal for use with adults. Although Wonnacott et al.'s (2008) experiments using the verb-argument structure paradigm also used "minimal exposure" items, their items were more incidental as they occurred in a more complex context, alongside some of the input items as well as other novel items. In the current experiment, by contrast, learners heard the two *minimal-exposure nouns* in a block of eight sentences with four sentences per noun without any other *input nouns* included. As the production test immediately followed, the restrictions on these nouns were likely to be very salient for adults (in fact, 75% of adults scored 100% in the minimal-exposure nouns test). Hearing the new nouns in a separate block may also have encouraged adults to feel that these items should be treated as separate from other *input nouns*. This highlights that the same testing procedures may have different pragmatic considerations for adults and children.

General discussion

Two experiments explored the factors affecting generalization in language learning using an artificial language learning method-

ology. The artificial languages were designed such that nouns were followed by one of two meaningless particles, but some nouns were restricted to occur with only one of these particles. Using a noun with a non-attested particle could thus be viewed as an instance of overgeneralization on the part of the participant. We focused on the learning of noun-particle pairings which were matched in frequency across different input languages. Despite being matched for frequency, the extent to which pairings were reproduced versus generalized to the alternative particle depended upon the ambient language experienced.

Children (Experiment 1) and adults (Experiment 2) showed better learning of noun-particle dependencies under conditions of skew. For both groups, this effect was clearly seen in productions with nouns that had been included in the exposure set. In children, we also found evidence that experience with a *skewed* language conferred an ongoing advantage that transferred to the learning of two new nouns under conditions of minimal exposure: children previously exposed to *skewed* input learned minimal-exposure pairings better than children who had not. Adults did not show a reliable effect of skew for the *minimal-exposure nouns.* However, Bayes Factor analyses suggested that this test was insensitive. As discussed above (Experiment 2, Discussion), we feel that there are good reasons why this particular test may have been sub-optimal for adult participants.

As pointed out above, it is important to realize that our result here *cannot* be due solely to participants having stronger performance with the majority-particle nouns as a result of a bias to produce the more frequent particle. Participants are indeed predicted to show stronger learning for those nouns, but note that, to the same extent particle frequency benefits the majority-particle nouns, it should *weaken* the minority-particle noun. For this reason, it is important that both of these noun-types were included in our analyses. It is also important to take this result in conjunction with the findings of Wonnacott (2011), where there was also a benefit of previous exposure to a skewed-lexicalist input language seen with minimal-exposure items, although in this case the alternative *generalist* language was also equally skewed. Taken together, these results suggest that children's learning of lexically based patterns is stronger when the input distribution is skewed, and this initial strong lexically based learning can support ongoing learning of lexically based patterns in the input.

The ongoing benefit of skew which we see in these experiments fits with the general idea that higher-level learning about the more general nature of a language can affect the learning of lexically-specific patterns. The hierarchical Bayesian model used to model earlier data (Wonnacott, 2011; Wonnacott et al., 2008) has thus far not been used to make predictions about the consequence of skew. However, it does make use of a prior, which favors more *skewed* distributions. This prior – the so called Chinese Restaurant Prior – is commonly used in models where it is necessary to assign objects to classes where the number of classes is unknown (other priors which achieve this have the same bias for skew). If the number of classes is unknown, the model must always assume that there is some probability that a new object will be in a different class, though this probability decreases as data is sampled. This leads to a bias which favors distributions where some classes are more rare. Given the complexity of the hierarchical Bayesian

model, predictions with the languages used in the current experiments are unclear. Nevertheless, the use of this prior is at least consistent with an effect of skew.

The effect of skew may additionally be captured by the closely related *simplicity framework* which has also been used to model the process of constraining generalization (Hsu & Chater, 2010). This approach is able to capture the intuition that a grammar comprising a default rule and an exception is somehow simpler than one listing individual pairings for each noun. As a consequence, it should be easier to learn. Within the *simplicity framework,* and using the *minimal description length* principle (Hsu & Chater, 2010), the probability of acquiring a particular grammar (the relative difficulty in learning it) given set of utterances depends on both the "cost" of encoding the rules of the grammar (simpler grammars have lower encoding costs) and the "cost" of encoding the observed utterances under the grammar, with more accurately specified grammars benefiting from lower encoding costs. Thus, there is a trade-off between simpler grammars, which are low cost but incur higher encoding costs per utterance if they are less accurate, versus more complex grammars which have a higher cost but may be more accurate, and thus accrue lower encoding costs per utterance. A particular grammar is "acquired" when the overall cost (i.e., including both the encoding of the rules and utterances) is minimal. This means that as the number of utterances increases, it eventually becomes worthwhile to adopt a more complex yet accurate grammar (note that this can also be expressed in terms of a Bayesian model with a particular class of priors: the cost of the grammar is the prior, the cost of encoding the utterances the likelihood). Given sufficient input, learners can arrive at any grammar, but the quantity of input required will differ depending on the complexity of the grammar, with more complex grammars requiring more input to be cost-effective. This approach has been shown to predict order of acquisition and grammaticality ratings for natural languages in children (Hsu & Chater, 2010) and adults (Hsu, Chater, & Vitányi, 2011).

Returning to our experiments, the simplicity approach predicts that participants would begin with the most simple and fully general grammar, where nouns in general are followed by particles in general with no lexical specification. However, this overgeneral grammar will incur a cost when encountering both the *lexicalist-skewed* and the *lexicalist-unskewed* input sets, and this cost accumulates for each utterance encountered; this can be thought of as the model making erroneous "predictions" for particle occurrences that never occur. With time, disregarding the overgeneral grammar in favor of representing lexical patterns will become more cost effective for learning both languages. In addition, however, the simplicity metric predicts faster learning of lexical patterns given *skewed* than *unskewed* input: the *skewed* grammar is simpler since it is more efficient to have a default rule and an exception, rather than to list individual pairings for each noun; because it is simpler, this grammar should be arrived at with less input. The fact that children in the skewed language go on to show better learning for the minimal-exposure nouns suggests that stronger lexical learning for the input nouns continues to impact on their attention to lexically based information: they approach new lexically specific information (the noun-particle co-occurrences) in the context of a grammatical system which has moved further from the over-general grammar, and thus are more focused on learning these co-occurrences than children in the unskewed condition.

In addition to showing a benefit of skew, adults learned noun-particle relations better in the context of an entirely lexical language (i.e. one without alternating nouns). This fits with the results of previous work with adults (Perek & Goldberg, 2015; Perfors et al., 2010; Wonnacott et al., 2008). A hierarchical Bayesian model can capture this type of learning since the model forms an "overhy-

pothesis" about the extent to which particle usage is lexically determined across the language. This "overhypothesis" affects the likelihood of learning further noun-particle pairings. The perspective of the simplicity approach is that following *lexicalist-unskewed* input, the relative cost per utterance is much steeper for an overgeneral grammar than a lexically specified grammar, and thus savings per utterance encoding quickly accumulate. In contrast, in the *mixed* language, input from the alternating nouns provides support for this overgeneral grammar. Thus the *mixed* language is learned more slowly as more experience is required to move away from that grammar.

Unlike for adults, there was no evidence that children in Experiment 1 benefitted from overall lexicality in the absence of skew. In fact, for the input nouns, there was evidence that learning of the noun-particle relationships was equivalent in the *lexicalist-unskewed* and *mixed* language (BF < 1/3, supporting the null hypothesis), suggesting that experience of alternation had no effect on their learning. The results from the minimal-exposure items were inconclusive, with the BF analyses suggesting H1 could not be either accepted or rejected. It is thus necessary to be cautious, although, on the basis that the effect for skew is bigger for the input nouns than the minimal-exposure items, we think it unlikely that experience of lexicality would benefit these items but not the nouns actually occurring in the input. Given this, we tentatively suggest that the stronger learning observed by Wonnacott (2011) in the *lexicalist* language compared with the *generalist* language was dependent on the fact that the skewed distribution of particles supported strong learning of lexical patterns in the *lexicalist* condition.

If overall input lexicality does not confer a benefit for children in the absence of skew, can this be accommodated by approaches such as a hierarchical Bayesian model or the simplicity framework? First, it is important to note that the current results, taken in conjunction to those of Wonnacott (2011), do not speak against the general claim that that strong learning of some lexical restrictions can boost the further learning of others (a key component to the notion of "over-hypothesis" in the hierarchical Bayesian model). As discussed above, our results sit best with an account in which exposure to skewed, lexically based input can lead to strong learning of lexical constraints which aids the learning of further input constraints. On the other hand, we do not see evidence that children can benefit from lexicality if there is no skew to aid their learning.

One possibility is that children have a stronger "prior" working against the learning of a lexically-specified grammar, arising from the greater capacity required for storing such a grammar compared with one in which particle usage is fully generalized. Apparently against this proposal is evidence that children's usage of particular structures is often highly item specific in the early stages of learning, which has been argued to provide evidence of more lexically conservative learning (e.g. Lieven, Pine, & Baldwin, 1997; Tomasello, 2000; Wonnacott et al., 2012). Our view is that the extent to which early learning is lexically specific likely depends on the nature and complexity of the structures under discussion. In the current experiment, we use the simplest possible "structure" (a single word), making generalization relatively easy, and our learning task exacerbates pressures on memory by asking children to learn and reproduce multiple lexical associations in tandem. These factors may bias children towards generalization.

Regarding the fact that children do not show a learning advantage in the *lexicalist-unskewed* compared to the *mixed* condition: it is possible that they have not yet sufficiently mastered the lexical nature of the language for overall lexicality to aid learning. That is, if we were to provide more input, at some point in learning, learning the restrictions on some *input nouns* would in fact begin to confer an advantage on others, and would also confer an ongoing

advantage for the learning of the *minimal-exposure nouns*, in line with the predictions of the models. To fully explore this, it would be necessary to observe the learning of these languages over a longer time frame, observing multiple time points so that we can watch participants "retreat" from overgeneralization with input nouns. This would reveal whether this process occurs more quickly in the lexicalist language than in the mixed language, and whether at some point children start to show the predicted differences in performance with minimal exposure nouns. (Note that experiments looking at learning over a longer time frame, although practically challenging, could potentially also inform the claim that there is a sensitive period for second language learning, such that successful acquisition is less likely after this period (Johnson & Newport, 1989; DeKeyser, 2012). One possibility is that older learners are less likely to fully retreat from over-generalization, leading to the proposal that it might be possible to induce situations where older learners remain "stuck" on an overgeneralized grammar. Experimental designs using artificial languages offer a way to tease apart age from factors that are inherently correlated with age in natural language learning situations.)

Returning to the fact that children do not show the predicted lexicality benefit in the current experiments: as discussed above, the computational approaches we have described predict that eventually, given sufficient exposure, lexically based patterns for input nouns would be fully encoded in all languages, however this will happen more quickly for in the lexicalist language than in the mixed language. The alternative, however, is that there may be no circumstance in which children will show a benefit of overall lexicality suggesting that the approach embodied in the hierarchical Bayesian and simplicity models may not be relevant to child language learning, or at least not child learning as captured in Experiment 1. For example, it is possible that the benefit to adults occurs as a result of more "top down", strategic approach. Future research is needed to explore the predictions of these approaches and to confirm the circumstances in which adult and child learners are sensitive to overall lexicality. Another question of interest is whether the patterns of higher level learning captured in computational models such as the hierarchical Bayesian and simplicity models, if they are indeed relevant to children's learning, could arise from lower level more mechanistic accounts of learning such as a discrimination learning model (Ramscar & Baayen, 2013; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010).

The relationship between explicit awareness and learning in these types of statistical learning experiments is not well understood. As a first step to probing this relationship, we assessed whether participants were explicitly aware that particle usage could be conditioned on the noun and related this to learning in the different conditions. To be coded as "aware", participants had to directly mention one (or more) of the noun-particle associations in their input or (as only ever for adults) to make some more general comment indicating that they realized particle usage was lexically conditioned (for example, describing particle usage in terms of animal "gender"). Not surprisingly, adults showed greater awareness across all conditions than children. However, for both groups, awareness differed by condition. For adults, we saw more awareness in both the *lexicalist-unskewed* and *lexicalist-skewed* conditions than in the *mixed* condition, whereas for children there was greater awareness in the *lexicalist-skewed* condition than any of the other conditions (although not reliably so). Thus the conditions where participants showed greater awareness were generally those where learning of the noun-particle relationships was strongest (though adults were not more aware in the *lexicalist-skewed* than the *lexicalist-unskewed* condition, despite stronger learning in the former). Interestingly, however, within each condition, participants showing greater awareness did not show better learning. Thus, there is no clear evidence that awareness drove stronger learning.

One possibility for why the number of "aware" participants differed across the conditions may be that the input languages are more or less difficult to articulate, especially given the main index of awareness was recall of a noun-particle relationship. For example, in the *lexicalist-skewed* language, children might have found it easier to recall and articulate the "exception animal". For adults, in the *lexicalist-unskewed* languages, there were more fixed noun particle relationships (eight; six *input*, two *minimal-exposure*) than in the *mixed* language (four; two *input*, two *minimal-exposure*) from which they could recall any one when questioned. Alternatively, it may be that, for example, exposure to the skewed language does specifically lead to explicit awareness, or semi-awareness, that the language had "rules" and an "exception", even for children. To some extent this may also be the case in natural languages. For example, English children probably have some awareness that plurals are marked with an -*s* but could also list some exceptions. It is important to note that these are speculations. In general it is difficult to know whether participants who do not articulate the patterns might not nevertheless be aware of them if questioned appropriately. For example, it is possible that participants (particularly adults) may have thought the idea that particles might be associated with particular animals was too obvious to mention. We continue to collect similar data in ongoing experiments, probing awareness in different ways so as to address the relationship between awareness and learning, as well as the question of why some individuals should show more awareness than others.

In summary, our findings add to a body of work which suggests that purely formal, distributional statistics may play a role in grammar learning (cf., Elman, 1998; Mintz, 2002; Reeder, Newport, & Aslin, 2013; Wonnacott et al., 2012). This does not negate there being a role for other types of cues in learning; cues from phonology and semantics have been shown to be important for example (Ambridge, Pine, & Rowland, 2012; Ambridge et al., 2014; Culbertson, Gagliardi, & Smith, 2017; Fitneva, Christiansen, & Monaghan, 2009; Perek & Goldberg, 2015). Artificial language learning provides an ideal methodology for exploring the interplay between different cues, and this is something that we are currently exploring in other work.

Given that the particular distributional cue of skew does appear to modulate the learnability of arbitrary, lexically based patterns, it is interesting to investigate the extent to which this is generally reflected in distributions found in natural languages. It is notable that grammatical systems are often described in terms of majority forms, which function as regular/default "rules" (e.g. English plural –*s*), along with exceptions (e.g., *feet*, *children*); the current work suggests this may make these systems more learnable. Another place where it could be interesting to explore a role for skew is in grammatical gender systems. Gender systems apparently require extensive lexically based learning, since nouns are assigned to different gender-classes in a semi-arbitrary way. The necessary corpus study is beyond the scope of the current work, yet it is interesting to note that linguists have long assumed that languages have a "default" gender (for example, masculine gender in the Romance Languages) which is assigned in the absence of other cues to gender, and is assumed to be the one that occurs with the majority of nouns in the language. Moreover there is some limited evidence that these "defaults" affect noun assignment. An elicited production study with French children reported a tendency to assign the "default" masculine gender (Boloh & Ibernon, 2013; although there is an interesting question as to how this "default" assignment interacts with phonological and semantic cues, e.g., Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981; Mulford, 1985). An important question is just how *skewed* a language needs to be to create a "default" and incur a learning advantage. Our experiments used a language with a 1:5 skew, but would a 2:5 ratio still be sufficient? It is also possible that a single exception may have a

special psychological status, so that a 1:5 and 2:10 skew might not be equivalent. These are questions to be explored in future corpus and experimental work. For now, we can conclude that learning by both children and adults is affected by the higher level distributional properties of the input, that is, above the bigram level, and that successfully learning the restrictions on some nouns can aid learning of the restrictions on others. This rules out any account of learning which relies only on frequency at the lexical level.

Acknowledgments

References

Alloway, T. P. (2008). *Automated working memory assessment*. London: Pearson Assessment.

Ambridge, B., Pine, J. M., & Rowland, C. F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition, 123*, 260–279. http://dx.doi.org/10.1016/j.cognition.2012.01.002.

Ambridge, B., Pine, J. M., Rowland, C. F., Freudenthal, D., & Chang, F. (2014). Avoiding dative overgeneralization errors: Semantics, statistics or both? *Language, Cognition and Neuroscience, 29*, 218–243. http://dx.doi.org/10.1080/01690965.2012.738300.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry, 10*, 533–581.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear Mixed Effects Models using S4 Classes. R package version 0.999999-2 Retrieved from <http://cran.rproject.org/web/packages/lme4/index.html>.

Boloh, Y., & Ibernon, L. (2013). Natural gender, phonological cues and the default grammatical gender in French children. *First Language, 33*, 449–468. http://dx.doi.org/10.1177/0142723713499848.

Braine, M. D. S. (1971). On two types of models of the internalization of grammars. In D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium* (pp. 153–186). New York: Academic Press.

Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76–95. http://dx.doi.org/10.1006/jmla.1993.1005.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping of a phrasal form and meaning. *Developmental Science, 8*, 500–508. http://dx.doi.org/10.1111/j.1467-7687.2005.00441.x.

Culbertson, J., Gagliardi, A., & Smith, K. (2017). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language, 92*, 343–358. http://dx.doi.org/10.1016/j.jml.2016.08.001.

DeKeyser, R. M. (2012). Age effects in second language learning. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition*. London, UK: Routledge.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781. http://dx.doi.org/10.3389/fpsyg.2014.0078.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72*, 78–89. http://dx.doi.org/10.1016/j.jmp.2015.10.003.

Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. In M. Overgaard (Ed.), *Behavioural methods in consciousness research* (pp. 199–220). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199688890.003.0012.

Elman, J. L. (1998). Generalization, simple recurrent networks, and the emergence of structure. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the twentieth annual conference of the cognitive science society*. Mahwah, NJ: Erlbaum.

Fitneva, S., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language, 36*, 967–997. http://dx.doi.org/10.1017/S0305000908009252.

Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language, 90*, 58–89. http://dx.doi.org/10.1353/lan.2014.0013.

Goldberg, A. E., Casenhiser, D., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15*, 289–316. http://dx.doi.org/10.1515/cogl.2004.011.

Hsu, A., & Chater, N. (2010). The logical problem of language acquisition goes probabilistic: No negative evidence as a window on language acquisition. *Cognitive Science, 34*, 972–1016. http://dx.doi.org/10.1111/j.1551-6709.2010.01117.x.

Hsu, A., Chater, N., & Vitányi, P. (2011). The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Cognition, 120*, 380–390. http://dx.doi.org/10.1016/j.cognition.2011.02.013.

Hudson Kam, C. L., & 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. http://dx.doi.org/10.1207/s15473341lld0102_3.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*, 30–66. http://dx.doi.org/10.1016/j.cogpsych.2009.01.001.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99. http://dx.doi.org/10.1016/0010-0285(89)90003-0.

Karmiloff-Smith, A. (1981). *A functional approach to child language: A study of determiners and reference* (Vol. 24) New York, NY: Cambridge University Press.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*, 307–321. http://dx.doi.org/10.1111/j.1467-7687.2007.00585.x.

Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24*, 187–219. http://dx.doi.org/10.1017/S0305000996002930.

Matthews, D., & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive Science, 34*, 465–488. http://dx.doi.org/10.1111/j.1551-6709.2009.01091.x.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*, 678–686. http://dx.doi.org/10.3758/BF03196424.

Mulford, R. (1985). Comprehension of Icelandic pronoun gender: Semantic versus formal factors. *Journal of Child Language, 12*, 443–453.

Perek, F., & Goldberg, A. (2015). Generalizing beyond the input: The functions of the constructions matter. *Journal of Memory and Language, 84*, 108–127. http://dx.doi.org/10.1016/j.jml.2015.04.006.

Perfors, A., Tenenbaum, J., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language, 37*, 607–642. http://dx.doi.org/10.1017/S0305000910000012.

Perfors, A., & Wonnacott, E. (2011). Bayesian modeling of sources of constraint in language acquisition. In I. Arnon & E. V. Clarke (Eds.). *Experience, variation and generalization: Learning a first language* (pp. 277–294). Philadelphia: John Benjamins. http://dx.doi.org/10.1075/tilar.7.16per.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Powell, M. J. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. Cambridge.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*, 413–425. http://dx.doi.org/10.1016/j.jml.2008.02.002.

Ramscar, M., & Baayen, R. H. (2013). Production, comprehension and synthesis: A communicative perspective on language. *Frontiers in Language Sciences, 4*, 233. http://dx.doi.org/10.3389/fpsyg.2013.00233.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science, 34*(6), 909–957. http://dx.doi.org/10.1111/j.1551-6709.2009.01092.x.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology, 66*, 30–54. http://dx.doi.org/10.1016/j.cogpsych.2012.09.001.

Rouder, J. N. (2014). Optional stopping: No problem for bayesians. *Psychonomic Bulletin & Review, 21*, 301–308. http://dx.doi.org/10.3758/s13423-014-0595-4.

Samara, A., Smith, K., Brown, H., & Wonnacott, E. (in press). Acquiring variation in an artificial language: Children and adults are sensitive to socially-conditioned linguistic variation. *Cognitive Psychology*.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209–253. http://dx.doi.org/10.1016/S0010-0277(99)00069-4.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language, 65*, 1–14. http://dx.doi.org/10.1016/j.jml.2011.03.001.

Wonnacott, E., & Perfors, A. (2009). Constraining generalisation in artificial language learning: children are rational too. In *Paper presented at the 22nd annual CUNY conference on human sentence processing*. Davis, California, USA.

Wonnacott, E., Boyd, J. K., Thomson, J. J., & Goldberg, A. E. (2012). Verb argument structure. Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language, 66*, 458–478. http://dx.doi.org/10.1016/j.jml.2011.11.004.

Wonnacott, E., Perfors, A., & Tenenbaum, J. B. (2008). Higher order inference in verb argument structure acquisition. In *Paper presented at the 14th annual conference on architectures and mechanisms for language processing*. Cambridge, UK.

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. http://dx.doi.org/10.1016/j.cogpsych.2007.04.002.