Memory limitations alone do not lead to over-regularization: An experimental and computational investigation

Amy Perfors (amy.perfors@adelaide.edu.au)
School of Psychology, University of Adelaide, Australia

Abstract

The "less is more" hypothesis suggests that one reason adults and children differ in their language acquisition abilities is that they also differ in other cognitive capacities. According to one version, children's relatively poor memory may make them more likely to over-regularize inconsistent input (Hudson Kam & Newport, 2005, 2009). This paper investigates this hypothesis experimentally and computationally. Experiments in which adults were placed under a high cognitive load during a language-learning task reveal that in adults, increased load does not result in increased over-regularization. A computational model offers a possible explanation for these results, demonstrating that over-regularization should occur only in the presence of memory limitations *as well as* a strong prior bias for over-regularization. Taken together, these findings suggest that the difference in over-regularization between adults and children may not be attributable solely to differences in memory capacity between the two groups.

Keywords: language acquisition; over-regularization; statistical learning; memory; computational modelling

Introduction

In many ways, ranging from phonetic perception to aspects of syntax, children are superior language learners than adults. Some argue that this is because language acquisition in children is guided by language-specific acquisition procedures, whereas adult acquisition is directed by more domain-general learning mechanisms (e.g., Bley-Vroman, 1990). However, there are many other possibilities, since children and adults also differ profoundly in their cognitive capabilities, knowledge, assumptions, and typical linguistic input. Learning a second language is made more difficult by interference from the first language (e.g., Mayberry, 1993; Iverson et al., 2003), adult brains are less malleable than the brains of children (Elman et al., 1996; MacWhinney, 2005), and adults and children differ in the nature of the social support (Snow, 1999) and linguistic input (Fernald & Simon, 1984) they receive.

One hypothesis, often called "less is more", suggests that the relative cognitive deficits in children may actually *help* with language acquisition, either by enabling them to isolate and analyze the separate components of a linguistic stimulus (Newport, 1990), or by leading them to over-regularize inconsistent input (Hudson Kam & Newport, 2005, 2009). Children do indeed over-regularize while adults do not. Deaf children exposed to the inconsistent sign language of hearing parents will over-regularize that language and produce regular grammatical forms (Singleton & Newport, 2004), as will children exposed to inconsistent input in an artificial language (Hudson Kam & Newport, 2005). By contrast, adult language learners are known to produce highly variable, inconsistent utterances, even after years of experience with the language and after their grammars have stabilized (Johnson, Shenkman, Newport, & Medin, 1996).

Although children's tendency toward over-regularization is well-established, the reason for the difference between adults and children is far from clear. The "less is more" hypothesis suggests that over-regularization may be due to some aspect of children's cognitive capacities, such as their poorer memory. Adults do tend to over-regularize more when the input is complex, when the probabilities involved are small (Hudson Kam & Newport, 2009), or when lexical retrieval is more difficult (Hudson Kam & Chang, 2009). This may be because more complex input imposes more of a load on their cognitive resources. The "less is more" hypothesis is also supported by computational work (Elman, 1993) suggesting that learning is easier when early input is simpler (although that work does not speak to the issue of over-regularization). In general, there has been little computational or experimental research that directly measures or manipulates memory or processing speed and evaluates whether these are associated with different degrees of over-regularization in adults.

In previous work I investigated whether adults placed under cognitive load over-regularized more (Perfors & Burns, 2010). The logic was that if deficiencies in the particular capacities involved in the load tasks are what cause children to over-regularize, then adults under heavy load should behave more like children in their pattern of over-regularization. Adults took part in a standard word-learning task, but in some conditions they also had to solve equations (OPERATIONAL LOAD) or read sentences (VERBAL LOAD) in between word-learning trials. People did not over-regularize in any condition, regardless of load. However, because the additional load tasks were interspersed rather than concurrent, it is possible that the load tasks did not interfere with memory enough to have an effect. In the first part of this paper I therefore report the results of an experiment that address this possibility by placing adults under *concurrent* memory load. As before, the participants failed to over-regularize in all conditions.

The second part of the paper uses a simple computational model to explain these results. The model systematically explores how different degrees and types of memory limitation affect over-regularization. It also investigates how memory limitations interact with prior biases for or against over-regularization. Results indicate that over-regularization only occurs when *both* memory limitations *and* a strong prior bias for over-regularization are present; neither alone is sufficient. Taken together with the experimental findings, they suggest that adult-child differences in over-regularization do not emerge from differences in memory capacity alone; adults may additionally have different prior biases about how to respond to inconsistent input.

Experiment

50 adults were recruited from the University of Adelaide and surrounding community and were paid $10. Subjects completed a word-learning task in which they were taught 10 novel two-word labels. Interspersed with the word-learning task was an interference task that required people to memorize a list of letters at the beginning of each trial and report that list at the end. Subjects in the HIGH LOAD condition had to memorize six letters each time, and in the LOW LOAD condition they had to memorize three. These results are compared to performance of an additional 25 subjects in a control NO LOAD condition, as reported in Perfors and Burns (2010).

The word learning task was modelled after a similar task described by Hudson Kam and Newport (2009). Their language contained 51 words, taught over the course of 9-12 days. Of critical interest was the evaluation of performance on the determiners, which were associated with nouns in an inconsistent fashion: participants heard the *main* determiner only 60% of the time. Conditions varied according to how many other determiners there were (always evenly distributed). Participants were asked to provide the noun and determiner associated with a scene and sentence and the frequency with which each determiner was produced was noted.

In order to remove extraneous elements of the task so as to focus on the determiner-production aspect, the "language" in this experiment consisted of 10 nouns, all two-syllable nonsense words mapped to images representing common objects. Each noun was followed by a one-syllable determiner: the *main* determiner occurred 60% of the time, and each of the four *noise* determiners occurred 10% of the time.[1] The specific image-label mapping and choice of *main* determiner was randomized for each participant.

The task consisted of a total of 200 trials of image-label pairs. On each trial, an image appeared on the computer screen and at the same time the person heard a female voice provide the label: for instance, they might see a picture of a baby and hear churbit mot. In the NO LOAD condition, participants went to the next trial by clicking a next button. In the two load conditions, each image was preceded by a list of letters to memorize (six letters in the HIGH LOAD condition and three in LOW LOAD), which was visible for 2.5 seconds. The image was displayed for 1.5 seconds and followed by a response phase in which participants reported the last set of letters in order. At that point memorization accuracy and time taken to respond were displayed, and when the participant pressed next the next set of letters to memorize appeared.

Learning was tested with 10 questions every 50 trials, for a total of 40 test questions. At each test, the participant was presented with an image and asked to verbally produce the label for it, which the experimenter (who was blind to the correct answers) wrote down. No feedback was given.

[1]Noun words used were: dragnip, raygler, churbit, tramdel, shelbin, pugbo, wolid, foutray, nipag, and yeetom. Objects used were: babies, balls, beds, birds, books, cars, cats, cups, dogs, and shoes. The five determiners were: mot, ped, sib, kag, and zuf.

Results

There are two natural questions we must answer.[2] First, is the load task difficult enough? Second, did participants in either of the load conditions over-regularize by producing the *main* determiner more than 60% of the time?

Was the load task difficult enough? There are two ways to evaluate whether the load tasks were sufficiently challenging to the cognitive capacities of the participants, whilst still being easy enough so that people could acquire at least some of the image-label mappings in the word-learning task. One indication that people were taking the load task seriously is that participants in both load conditions were reasonably accurate in memorizing letters: HIGH LOAD averaged 56% of letters correct (a mean of 3.4 letters per trial) and LOW LOAD averaged 85% correct (2.5 letters per trial).[3] Another indication is that participants learned fewer noun-image mappings if they were in the load conditions. Each person's answers were coded as *correct* if the noun they produced was identical to or phonologically similar to the correct noun for that image (e.g., wolin instead of wolid). Participants performed above chance in all conditions, but significantly worse in HIGH LOAD (41%) and LOW LOAD (47%) than in NO LOAD (67%), suggesting that the interference tasks were, indeed, imposing significant strain on their cognitive resources.[4]

Did adults over-regularize more when under cognitive load? Following Hudson Kam and Newport (2009), participants who did not get at least 9 out of the final 20 nouns correct on the test trials were excluded.[5] Then, on every valid trial (in which a correct noun was produced), I calculated the percentage of time either the *main* determiner, any other determiner (*noise*), or *no* determiner was produced. Figure 1 demonstrates that there were no significant differences between conditions in terms of *main* determiner production: that is, participants in the load conditions were not significantly more likely to over-regularize.[6]

[2]In Perfors and Burns (2010) we addressed a third question: whether lower performance on a separate working memory task predicted greater over-regularization in the word learning task. I performed a similar analysis here and found no evidence for such a relationship, but do not have space to describe this analysis in detail.

[3]To ensure that results were not due solely to participants who did not take the load task seriously, all analyses were repeated after excluding participants who got fewer than 50% or 70% correct; in both cases, all results are qualitatively identical.

[4]A one-way Anova on nouns correct by condition was significant: $F(2,72) = 7.56, p = 0.001$. Post-hoc comparisons using the Tukey-Kremar test indicated that the mean score for the NO LOAD condition (M=0.667, SD = 0.05) was significantly different than the mean for the HIGH LOAD ($M = 0.407, SD = 0.05$) and LOW LOAD ($M = 0.473, S = 0.05$) conditions, but the latter two were not significantly different from each other.

[5]This resulted in keeping 23 subjects in the NO LOAD condition, 15 in HIGH LOAD, and 19 in LOW LOAD. All analyses were also performed without this exclusion; results were qualitatively identical.

[6]One-way Anova on main determiner production by condition: $F(2,49) = 2.05, p = 0.1393$. To further explore this outcome, a post-hoc comparison using Tukey-Kramer indicated no significant difference between any of the conditions compared pairwise.

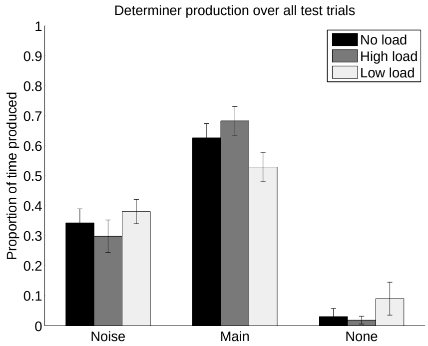

Figure 1: Performance by condition in determiner production. There was no significant difference between conditions in tendency to over-regularize. Error bars are standard error.

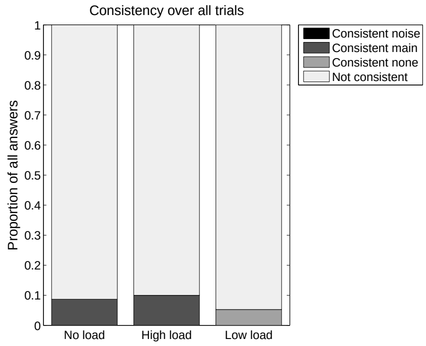

Figure 2: Individual consistency in determiner production by condition. For the most part, few participants showed any consistency in their pattern of determiner usage, and those in the load conditions did not tend to be more consistent.

This is suggestive, but because it is an analysis of mean performances this outcome may be hiding individual over-regularization in different directions. To evaluate this possibility, following Hudson Kam and Newport (2009) a "consistency threshold" of 90% was set: each participant was coded as *consistent main*, *consistent noise*, or *consistent none* if they produced the determiner type in question on at least 90% of the valid trials, and *not consistent* if they did not.[7] Figure 2 shows that few participants were consistent in any way, and differences between conditions were minor. In order to determine if the tendency to over-regularize changed as they acquired more of the language, analyses for both Figures 1 and 2 were repeated for the first and second half of testing. Results were qualitatively similar for all analyses.

Hudson Kam and Newport (2005, 2009) hypothesize that differences in cognitive capabilities between children and adults may lead to differences in regularization, either because children are less efficient at encoding memories, or because they have more difficulty retrieving specific memory forms. Either way, the theory suggests that children will over-regularize some forms and fail to produce others, but adults will store and retrieve the memories more veridically.

Another possibility is that children simply have a prior bias to favor regularization, whereas adults do not. This bias might be language-specific (e.g., Bickerton, 1984) or more domain-general; either way, it would result from something other than age-related differences in memory capacity. In the next section I use a computational model to investigate the expected effects of both prior biases and memory limitations, and how they trade off against each other.

Computational analysis

Most tasks in which there is the potential for over-regularization can be described abstractly as tasks in which there are k possible outcomes and the learner must learn the distribution over those outcomes. In this experiment there are six outcomes associated with each noun (five for each of the determiners, and one for no determiner), while in a typical probability matching task, the outcomes might be the frequency of different colors of flashing lights or cards in a deck.

This situation is captured by the multinomial distribution, where θ_i denotes the probability of outcome i and $\sum_{i=1}^{k} \theta_i = 1$. In a multinomial, the data for the observed outcomes $\mathbf{y}$ are generated from the underlying θ according to:

$$p(\mathbf{y}|\theta) = \binom{k}{y_1 \ldots y_k} \prod_{i=1}^{k} \theta_i^{y_i}. \quad (1)$$

The task of the learner is to reason backward from the outcomes $\mathbf{y}$ to infer the nature of the underlying "true" distribution θ. Which distribution is learned will depend on two things: the nature of the data $\mathbf{y}$ and any prior beliefs about what θ should look like.[8] A natural, mathematically elegant, and widely used prior for multinomial data is the Dirichlet distribution (Gelman, Carlin, Stern, & Rubin, 2003). This model uses a symmetric Dirichlet distribution, which imposes no prior bias in favor of any one outcome more than another across the whole dataset. Symmetric Dirichlet distributions have one parameter, α, which captures the degree to which each item (noun) is expected to be associated with only one outcome (determiner); it governs the extent of the bias for over-regularization. If α is very small, there is a strong bias for over-regularization: the model will assume that each noun is associated with only one determiner (although, because the prior is symmetric, it will have no prior bias about *which* determiner is most likely). When $\alpha = 1$, there is no bias for over-regularization; it is assumed that each outcome will occur with equal probability. I evaluate the role of the prior by considering four values of α: 1 (NO BIAS), 0.5 (WEAK BIAS), 0.05 (MEDIUM BIAS), and 0.005 (STRONG BIAS).

In addition to varying the strength of the prior bias for over-regularization, it is necessary to also model the effects of memory. How to do this is less obvious, but the most straightforward possibilities are to capture memory limitations via a corruption of the observed data $\mathbf{y}$ (which, in the uncorrupted

[7]The results do not change if the threshold is 70% or 80%.

[8]A complete absence of prior belief would mean that θ should always match the observed distribution $\mathbf{y}$ precisely; such a learner would never generalize beyond the input at all. It is possible to have very mild prior beliefs – e.g., the weak expectation that any outcome is equally likely – which would still enable some generalization.

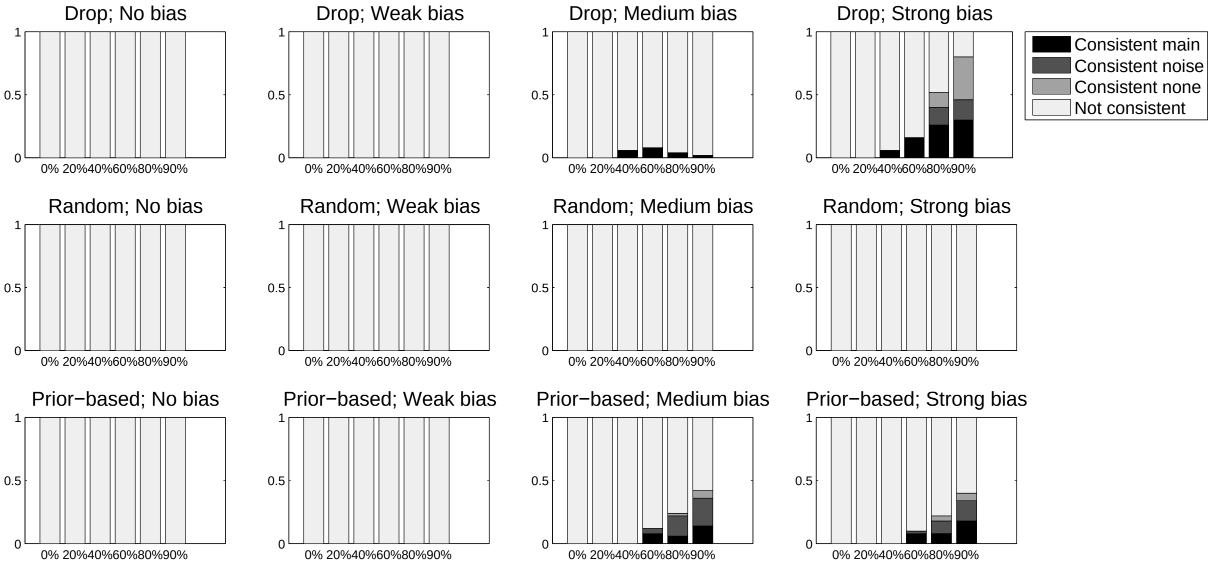

Figure 3: Model performance varying the strength of the prior bias (columns) and the effect of different kinds of memory limitation (rows). Each graph shows the proportion of *consistent* classifications out of 50 iterations (on the *y* axis) as a function of the percentage of memory affected (on the *x* axis): *n*% means that *n*% of the data are either dropped (DROP), flipped randomly (RANDOM), or reconstructed based on the prior (PRIOR-BASED). Over-regularization only occurs when memory is limited *and* there is a medium-to-strong prior for over-regularization.

case, always precisely follows the proportions of the input in the experiment: one determiner occurs 60% of the time, and four others occur 10% of the time).

How might memory corrupt the data? One possibility is to assume, as a first approximation, that memory loss simply means dropping data at random (the DROP condition). Dropping different proportions of the data would therefore map onto differences in memory capacity. Another possibility is to assume that memory limitations result in data being forgotten and then reconstructed by the mind. A trivial way to reconstruct such data would be to randomly randomly reassign "forgotten" data to any of the possible determiners with equal probability; this is the RANDOM condition. A more natural reconstruction method might be to presume that forgotten data is reconstructed according to the prior probability (the PRIOR-BASED condition). This can be modelled using the Chinese Restaurant Process:

$$P(\text{determiner } i|\text{previous data}) = \frac{n_i}{N+\alpha}$$
$$P(\text{new determiner}|\text{previous data}) = \frac{\alpha}{N+\alpha}$$

where n_i refers to the number of observations involving determiner i made so far, N is the number of observations total, and α is the same parameter that captures the prior bias. In fact, the Chinese Restaurant Process gives the same distribution as draws from a Dirichlet process, which is why it is a natural way to capture memory loss within this model.

Predictions about the expected distribution over outcomes given the data and the priors are given by Bayes Rule:[9]

[9]The integral in the denominator is calculated by drawing 10,000 samples of θ from the Dirichlet distribution parameterized by α.

$$P(\theta|\mathbf{y},\alpha) = \frac{P(\mathbf{y}|\theta,\alpha)P(\theta|\alpha)}{\int_{\theta'} P(\mathbf{y}|\theta',\alpha)P(\theta'|\alpha)d\theta'} \quad (2)$$

Figure 3 shows expected performance by prior bias and memory. To make the model results comparable to the experimental findings, consistency is calculated the same way as in the experiment: e.g., *consistent main* means that on that iteration the model predicted that 90% or more of the determiners should be the *main* one. Each of the stacked bars reflects the proportion of runs (out of 50) in which the model achieved any of the kinds of consistency.

There are two striking things about these results. First, they demonstrate that simply having a prior bias for over-regularization is insufficient to cause over-regularization. This is because the quantity of data must also be small. In this model, memory limitations had the effect of limiting the quantity of (accurate) data, but other data-limiting factors might also include bottlenecks in the input or attentional restrictions. The reason that a prior bias alone is insufficient is because a sufficient quantity of data will always overcome any prior; a rational learner should think it much more likely that a given determiner actually occurs 60% of the time if it is observed in 600 out of 1000 observations rather than 3 out of 5. Because quantity of the data matters, a prior bias only has an effect when there is little veridical data available.

The second implication of these results is that memory limitations alone do not result in over-regularization either. Memory limitations must occur along with some sort of prior bias for over-regularization, whether it comes in when memory is being reconstructed or when interpreting situations where there are few observations. The reason a prior bias is necessary is because without it, memory limitations don't change the overall *pattern* of data. For instance, suppose a learner was exposed to 10 determiners following the distribution in the data (6-1-1-1-1-0). If the learner randomly forgot 60% of them, they would be unlikely to forget all of the *noise* determiners and more likely to forget some of each type. Without a strong prior towards over-regularization, the learner wouldn't ignore the *noise* items that remain, and thus would not over-regularize. Even in the extreme where 90% of the data is forgotten, over-regularization should not occur: when there is very little data the prior is weighted more heavily, so without a prior for over-regularization, a learner given very little data will assume that any outcome is possible.

Discussion

The "less is more" hypothesis suggests that one reason for the difference between adult and children in language acquisition is due to unequal cognitive capacities: children's poor memory may make them more likely to over-regularize inconsistent input. In an experiment building on Perfors and Burns (2010), adults were placed under a high cognitive load and the effect of this manipulation was evaluated. Although the cognitive load was strong enough to impair performance, increased load did not lead to increased over-regularization. Modeling work demonstrates that over-regularization should only emerge if the learner is has both a limited memory and a strong prior bias for over-regularization.

Taken together, these results suggest that memory limitations are necessary but not sufficient for over-regularization to emerge, and therefore memory differences between children and adults cannot be the only reason children but not adults over-regularize. This finding is consistent with other work showing that children with better memories or faster processing speed actually do *better* at learning language (e.g., Fernald, Perfors, & Marchman, 2006; Rose, Feldman, & Jankowski, 2009). It maybe that children do have some sort of prior bias favoring over-regularization that adults lack, but it is worth considering possible limitations first.

It is theoretically possible that the load tasks were not difficult enough to limit adults' memories to the point that any effects would be visible. However, this seems unlikely, for two reasons. First, the load tasks significantly impaired people's ability to learn the nouns, indicating that they placed a heavy burden on the learners. Second, even in analyses where one would expect poorer performance (e.g., on just the first half of test trials, or including even participants who learned very few nouns) there was no tendency toward over-regularization. This suggests that the reason adults failed to over-regularize was *not* that they simply found the task too easy. For similar reasons, it is unlikely that the simplicity of the task (learning noun-determiner pairings rather than full languages) is the reason for the findings; I will pursue this in future work.

On first glance, these findings might appear to contradict those of Hudson Kam and Chang (2009), who found that over-regularization in adults could be diminished by improving the ease of lexical retrieval. However, they aimed to make adults *less* like children by making the cognitive load easier, rather than to make adults act *more* like children by making it harder. It is possible that there is an inherent asymmetry to adults' performance: that it is relatively easy to make adults over-regularize less, but that getting them to regularize more is difficult. The computational model is consistent with this possibility, and such an asymmetry certainly exists in decision-making problems, in which great efforts have been made to stop adults from probability matching, to little avail (e.g., Shanks, Tunney, & McCarthy, 2002).

The computational model in this paper was deliberately chosen to be extremely simple in order to minimize the extent to which the results are dependent on arbitrary modeling choices. There was only free parameter in the model (α) and it was systematically varied. The multinomial distribution is the most obvious and widely-used way of capturing distributional data in which many outcomes are likely, and the Dirichlet distribution is the most widely-used and mathematically elegant prior for multinomial data. Memory limitations were modeled simplistically, but in a way that captures to a first approximation the different ways in which memory constraints might have an effect (either losing information or distorting it in different ways). Moreover, the qualitative results were driven by model-independent factors: a prior bias is necessary because memory limitations alone do not change the *pattern* of data remembered, and some sort of data-limiting mechanism (like a memory constraint) is necessary because otherwise any prior bias for over-regularization will be overwhelmed by the inconsistent data in the input. It is therefore unlikely that incorporating a more realistic memory model would change these results in any qualitative way, although this topic is being explored further in my lab.

One assumption inherent in the model is that it is Bayesian, meaning that it predicts the behavior of a rational learner. This means that the importance of previous biases (the prior) and fitting the data (likelihood) are balanced in a particular way (according to Bayes' Rule). However, every model needs to perform *some* tradeoff between these two factors. Because of this, models that weigh these tradeoffs differently might vary quantitatively, but all models except for the most pathological[10] should show that over-regularization is more likely when the input is limited and the prior bias for it is strong.

It is also worth noting that, although the model is Bayesian, this is not an ideal learning analysis; because the model incorporates different kinds of memory limitations, it should be more properly understood as a "capacity-limited" rational model. It thus allows us to investigate what a rational learner *with certain capacity constraints* might be expected to do. This sort of approach is an important step toward bridging computational-level and process-level accounts of cognition.

One simplification this model makes is that it is incapable of learning that different *kinds* of items might be associated with very different distributions (e.g., that some nouns are associated with only one determiner, but some are associated

[10]"Pathological" models include those that don't learn at all from data or never generalize at all beyond the data.

with many). That extra complexity was unnecessary to model this experiment, in which all items have the same distribution of determiners and consistency is calculated across the entire dataset. Existing models corresponding to an extension of this one have been used to capture complex phenomena including word-learning biases (Kemp, Perfors, & Tenenbaum, 2007; Perfors & Tenenbaum, 2009) and verb construction learning (Perfors, Tenenbaum, & Wonnacott, 2010).

How do we interpret the prior bias for over-regularization in the model? Independent evidence suggests that such a bias might be domain-general, since children over-regularize but adults do not even in non-linguistic domains like predicting which light will flash (e.g., Shanks et al., 2002; Weir, 1964; Derks & Paclisanu, 1967) or how often a cause will result in a given effect (Schulz & Sommerville, 2006). That said, the model incorporates no assumptions about the domain-specificity or domain-generality of the prior. It encodes the degree to which a bias for over-regularization exists, but the question of its origin is a matter for future work.

One final important point should be made: these findings are relevant only to the version of the "less is more" hypothesis that makes reference to over-regularization (Hudson Kam & Newport, 2005, 2009). The original theory is focused more on whether linguistic input is broken down into components or not (Newport, 1990); it suggests that as a result of their superior memories, adults may memorize entire frozen chunks of the input, while children – who are only able to recall smaller portions – find it easier to isolate linguistic components. This paper is not relevant to that version of the "less is more" hypothesis, since it postulates different mechanisms and applies to different phenomena.

It also remains possible that child-adult differences in over-regularization might be driven by cognitive factors that have effects beyond limiting or distorting input. Such differences could arise from variations in the ability to use metacognitive strategies (e.g., Flavell, Green, Flavell, Harris, & Astington, 1995). It may be that adults' ability to introspect and reason about their own cognition makes them more likely to rely on explicit rather than implicit learning (Ullman, 2004) – a difference that has been hypothesized to be the root of child-adult differences in language acquisition. A bias for over-regularization might result from a generalized preference for simplicity on the part of children. A great deal of work remains to be done to investigate the many possibilities that remain open, but this work suggests that memory alone is not the root of child-adult differences in the tendency to over-regularize inconsistent input.

Acknowledgments

Thank you to Natalie May, Jia Ong, Kym McCormick, and Tin Yim Chuk for their help recruiting participants and running the experiment.

References

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences, 7*, 173-221.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, *20*, 3–49.

Derks, P., & Paclisanu, M. (1967). Simple strategies in binary prediction by children and adults. *Jn. Exp. Psych.*, *73*(2), 278–285.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71–99.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fernald, A., Perfors, A., & Marchman, V. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Dev. Psych.*, *42*(1), 98–116.

Fernald, A., & Simon, T. (1984). Expanded information contours in mothers' speech to newborns. *Dev. Psych.*, *20*, 104–113.

Flavell, J., Green, F., Flavell, E., Harris, P., & Astington, J. W. (1995). Children's knowledge about thinking. *Monographs of the SRCD*, *60*(1).

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian data analysis* (2nd ed.). Chapman & Hall.

Hudson Kam, C., & Chang, A. (2009). Investigating the cause of language regularization in adults: Memory constraints or learning effects? *Jn. of Exp. Psych.: Lng., Mem., & Cog.*, *35*(3), 815–821.

Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang. Lng. & Dev.*, *1*(2), 151–195.

Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*, 30–66.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tokura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties with non-native phonemes. *Cognition*, *87*, B47–B57.

Johnson, J., Shenkman, K., Newport, E., & Medin, D. (1996). Indeterminacy in the grammar of adult language learners. *Journal of Memory and Language*, *35*, 335–352.

Kemp, C., Perfors, A., & Tenenbaum, J. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307-321.

MacWhinney, B. (2005). A unified model of language acquisition. In J. Kroll & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford Univ. Press.

Mayberry, R. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of american sign language. *Jn. of Speech and Hearing Res.*, *36*, 1258–1270.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11–28.

Perfors, A., & Burns, N. (2010). Adult language learners under cognitive load do not over-regularize like children. In *Proc. 32nd Annual Conf. of the Cognitive Science Society* (p. 2524-2529).

Perfors, A., & Tenenbaum, J. (2009). Learning to learn categories. In *Proc. 31st Annual Conf. of the Cognitive Science Society* (p. 136-141).

Perfors, A., Tenenbaum, J., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, *37*, 607-642.

Rose, S., Feldman, J., & Jankowski, J. (2009). A cognitive approach to the development of early language. *Ch. Dev.*, *80*(1), 134–150.

Schulz, L., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, *77*(2), 427–442.

Shanks, D., Tunney, R., & McCarthy, J. (2002). A re-examination of probability matching and rational choice. *Jn. of Behavioral Decision Making*, *15*, 233–250.

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

Snow, C. (1999). Social perspectives on the emergence of language. In B. MacWhinney (Ed.), *The emergence of language* (pp. 257–276). Mahwah, NJ: Lawrence Erlbaum Associates.

Ullman, M. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231–270.

Weir, M. (1964). Developmental changes in problem-solving strategies. *Psychological Review*, *71*, 473–490.