Probability matching vs over-regularization in language: Participant behavior depends on their interpretation of the task

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

Abstract

In a variety of domains, children have been observed to over-regularize inconsistent input, while adults are more likely to "probability match" to any inconsistency. Many explanations for this have been offered, usually relating to cognitive differences between children and adults. Here we explore an additional possibility: that differences in the social assumptions participants bring to the experiment can drive differences in over-regularization behavior. We explore this in the domain of language, where assumptions about error and communicative purpose might have a large effect. Indeed, we find that participants who experience less pressure to be "correct" and who have more reason to believe that any inconsistencies do not correspond to an underlying regularity do over-regularize more. Implications for language acquisition in children and adults are discussed.

Keywords: over-regularization; statistical learning; probability matching; language acquisition

Introduction

In a variety of situations, humans given probabilistic input will tend to *probability match* – that is, they respond differentially in a way that is proportional to those probabilities. In learning theory, this occurs when people (or animals) choose a stimulus proportional to the relative number of times it has been reinforced (e.g., Herrnstein, 1961, 1970; Baum, 1979; Pierce & Epling, 1983; Wearden, 1983). In decision making, this occurs when people are asked to predict the next item in a sequence (e.g., a card drawn from a deck, or a flashing light) and respond by choosing proportionally to the frequency of that item in the past (e.g., Castellan, 1974; Shanks, Tunney, & McCarthy, 2002; Vulkan, 2000). And in language learning, this occurs when people given linguistic input that varies inconsistently (such as an affix or particle occurring only 60% of the time, for no apparent reason) and they produce that particle proportional to its frequency in the input (e.g., Hudson Kam & Newport, 2005, 2009).

Although these cases vary widely in many details, what is interesting in all of them is that an overmatching or maximizing strategy may often be the more sensible one. This strategy, also called over-regularization in the language literature, involves producing or responding to the most frequent item closer to 100% of the time, rather than in a way that is proportional to its probability or frequency. Such a strategy is more sensible for different reasons in different domains. One receives more reinforcement if one always chooses the more frequently-reinforced stimulus; one makes more successful predictions if one always chooses the most frequent item; and one minimizes the burden on the listener as well as the chance of miscommunication by removing linguistic variability when it serves no purpose. Given this, why do people probability match?

One clue may come from the literature on children's behavior. Although children are less well-studied than that of adults, there is some evidence that preschool-aged children may overmatch or over-regularize more in a decision making or reinforcement learning context (Jones & Liverant, 1960; Derks & Paclisanu, 1967). This is consistent with the small amount of work on children in a linguistic domain suggesting that they are more likely than adults to over-regularize there as well (Hudson Kam & Newport, 2005, 2009). These findings are somewhat limited, since they are based on relatively few studies (especially in the area of language) and generally involve statistical rather than absolute differences: that is, more children over-regularize than adults, but some still do not. Nevertheless, they raise the intriguing possibility that whatever factor causes adults to probability match may play less of a role in preschool-aged children.

What might that factor be? A common hypothesis is that it is related to a cognitive change occurring between childhood and adulthood. One possibility is that preschool children have poorer metacognitive control – that is, they find it difficult to inhibit a previous response (Jones & Liverant, 1960; Weir, 1964), have a harder time monitoring and responding to conflict in their representations (Ramscar & Gitcho, 2007), and/or are more insensitive to the reward structure (Stevenson & Hoving, 1964). Another is that childrens' poorer memory and/or processing abilities might result in over-regularization (Hudson Kam & Newport, 2005, 2009).

Although there is support for many of these possibilities, it is also possible that the difference arises, at least in part, for a more prosaic reason: adults and children may be given slightly different tasks, or have different interpretations of the same task. It is natural for minor task differences to arise as a natural by-product of the effort to make the same experiment apply to widely varying age groups. There were indeed small differences in all of the experiments in which differential over-regularization is observed (e.g. Jones & Liverant, 1960; Derks & Paclisanu, 1967; Hudson Kam & Newport, 2005, 2009). However, we are more interested here in the kinds of differences in interpretation that might arise even with precisely the same methodology. That is, different groups (like children and adults) might have different interpretations or different assumptions that they bring to the *exact same* task. As explained below, this may be particularly an issue in linguistic tasks, which are the focus of this work.

This paper explores one main question: is human behavior in a language-learning domain affected by changes in the goal or assumptions underlying the task? It is known that adults can be pushed away from probability matching by varying the delay between changing a response and reinforcement (Baum, 1975), punishing people for some responses (Bradshaw, Szabadi, Bevan, & Ruddle, 1979), making one response easier (Bradshaw, Ruddle, & Szabadi, 1981), making rewards especially enticing (Shanks et al., 2002), or offering extensive corrective feedback (Shanks et al., 2002). However, none of these possibilities are relevant when the task is linguistic, since language learners generally receive little to no direct feedback or reinforcement, both in the real world (Pinker, 1989) and in this kind of task (e.g., Hudson Kam & Newport, 2005, 2009; Perfors & Burns, 2010; Perfors, 2011).

What goals or assumptions might adults and older children have that younger children do not, which cause the former to probability match but the latter to over-regularize? We hypothesize that adults and older children may feel a strong sense that there is a "right" answer, and a concomitant pressure to be pressure to be "correct" that younger children do not. We noticed in previous studies run in our lab (Perfors & Burns, 2010; Perfors, 2011) that adults reported a strong intuition that there must be some underlying reason behind the inconsistency in the linguistic input they received; for instance, one participant confided that they thought all of the "shiny things" were linguistically marked in the same way. In reality, there were no regularities between the inconsistent items and anything else in the experiment, but the participants did not know that and often tried to produce language in accordance with the regularities they thought they had observed.

Why did our adult subjects have such a strong intuition that there were regularities in the input? One possibility is that this is simply an intuition that all people typically bring to *any* language-learning scenario. It is not an unreasonable assumption; although some phenomena in language are truly arbitrary, much variation does occur for a reason. If people truly do always come to language learning tasks with this assumption, then we would expect adults to show the same behavior regardless of where they thought the language came from or what their goal in the learning task was. Conversely, this intuition may have been caused by, or at least exacerbated by, characteristics of the situation: being in an official lab, presented with stimuli that are clearly designed and non-accidental, and asked to learn about those stimuli all create the strong impression that there actually is some regularity there to learn. By contrast, real linguistic input contains many errors and inconsistencies that arise from the fact that it is produced on the fly, for communication, by real people. If young children are either more blind to the social pressures inherent in a lab-based experiment or more likely to interpret underlying irregularities in the input as errors, then this might be responsible for at least some of the observed difference in over-regularization between children and adults. After all, it is sensible to over-regularize inconsistencies if they do not hide some underlying regularity that you will be judged for missing.

A full test of our hypothesis would require us to manipulate children's beliefs about the nature of the input they are receiving and and experiment they are in. This is very difficult to do, and it is even more difficult to evaluate whether it has been done successfully. Alternatively, we can test the assumptions underlying our hypothesis by investigating adults. If adults respond to task characteristics that remove a pressure for generating "correct" responses by over-regularizing more, then this offers some support for the idea that at least part of the reason for the different behavior of children and adults might relate to different assumptions about the task.

Therefore, in this paper we explore the hypothesis that adult over-regularization behavior can be changed by changing the pressure for generating "correct" responses. This pressure is manipulated in two ways. First, we vary the cover story to change the assumptions people make about how likely the data is to reflect an underlying regularity of some sort. Second, we vary the goal of the task to emphasize or de-emphasize effective communication. Consistent with our hypothesis, we find that when the pressure for correct communication is reduced, adults over-regularize more. In the discussion we consider the implications of these results for experimental work on over-regularization, probability learning, and language learning more broadly.

Experiment

Participants. 52 adults[1] were recruited from the University of Adelaide and surrounding community and were paid $10 for their time. Participants were divided randomly into one of two conditions, HIGH PRESSURE and NO PRESSURE (described below). One person in the HIGH PRESSURE condition suffered a computer error causing a failure to save data, and two participants in the NO PRESSURE condition were excluded from the analysis for typing gibberish.[2] This left 25 participants in the HIGH PRESSURE condition and 24 in the NO PRESSURE condition.

Procedure. The standard task, which was the same in both conditions, involved a word learning task originally modelled after Hudson Kam and Newport (2009) and Perfors (2011). The original Hudson Kam and Newport (2009) taught a language containing many words taught over multiple days, but the key element for our purposes was that in this language, units (which they called determiners) covaried with the nouns in an inconsistent fashion: participants heard the **main** determiner only 60% of the time. Participants were asked to provide the noun and determiner associated with a scene and sentence and the frequency with which each determiner was produced after each noun was noted.

As in Perfors (2011), we removed extraneous elements of the task so as to focus on the aspect involved in producing the inconsistent units. Our language consists of words composed from 10 stems, all one-syllable consonant-vowel-consonant nonsense words mapped to images representing common ob-

[1]We ran 52 rather than the more round 50 because the HIGH PRESSURE condition required pairs of participants (an even number of people). We matched that number in the NO PRESSURE condition.

[2]This was reflected in their accuracy score, which was over four standard deviations below the mean accuracy for either condition.

jects. Each stem was attached to a by a one-syllable affix: the **main** affix occurred 60% of the time, and each of the four **noise** affixes occurred 10% of the time.[3] The main difference between this study and the previous ones is that while they were entirely auditory in both their presentation of stimuli and the modality of the response, in this study everything was written. This was necessary in order to make the experimental manipulation possible and believable, as explained below. Words were presented with no space between the stem and the affix: thus, participants would see words like PIMUT or JAFIG. They were not told that each of the words was composed from smaller units. The specific image-label mapping and choice of **main** affix was randomized for each participant.

As in Perfors (2011), 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 saw a label written in all capitals below it: for instance, they might see a picture of a baby and read YOKOM. People went to the next trial by clicking a next button. Learning was tested with 20 questions every 100 trials, for a total of 40 test questions. At each test, the participant was presented with an image and asked to enter the label for it. No feedback was given.

Conditions. The goal of this experiment was to explore the possibility that adult over-regularization behavior can be affected by changing the pressure for generating "correct" responses. We therefore constructed two conditions, one designed to increase this pressure as much as possible, and one designed to decrease it as much as possible.

HIGH PRESSURE. In this condition, we tried to increase the pressure to be correct by pairing each participant with another person who was in the lab at the same time. Each person was informed that the goal of this experiment was to learn a new language, and then successfully use it to communicate with the other person. They were asked to imagine they were scientists who had just discovered a community speaking this language, and they had gotten an informant to label a series of pictures for them. They were to read these labels, and then they would be tested on how well they had learned them by having to fill in the labels for new pictures. Participants sat at different computers and did the standard task individually, but at the end of the standard task each person was given the labels the other person generated during their test questions, and asked to match each of those labels with the correct image. The participants were told that they would get paid proportionally to how many of this final set of questions both of them got right. This created a great deal of social pressure to learn the language correctly, since not only was each individuals' payment dependent on it, so was their partner's.

It is important to note that this manipulation *in itself* does not favor either over-regularization or probability matching. Since the affixes did not correlate to the images at all, people could get 100% correct on the test regardless of what they did with the affixes, as long as the stems were matched to the correct image. Thus, any effect on over-regularization is due to increasing the sense that there is a "right answer" as well as the social pressure to find that right answer – *not* because participants could make more money using one strategy.

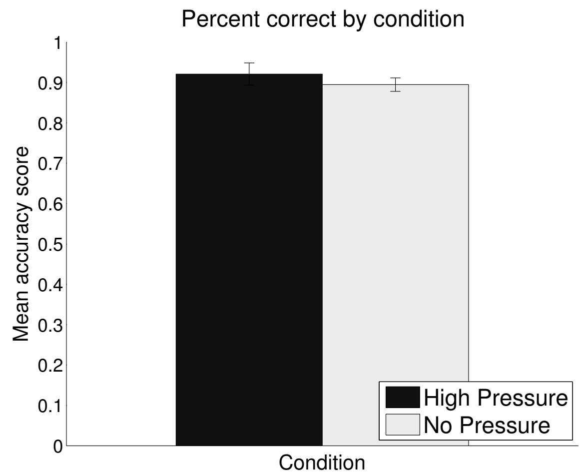

Figure 1: Accuracy in mapping the stems to the correct item. Both conditions showed high accuracy, and there is no significant difference between conditions. This suggests that any differences in over-regularization probably do not occur because participants were paying different amounts of attention to the stems and affixes.

NO PRESSURE. In this condition, the pressure for being "correct" was reduced in two ways: first, by not pairing participants with a partner and making their payment dependent on performance in any way; and second, by changing the cover story so that people were less likely to believe there was an absolutely correct answer. The cover story in this condition was that we were studying how languages change when multiple people learn it. Thus, a previous participant had learned some words in a fake language, and during the course of the learning they were presented with images and asked to label them with those words. Current participants were told that they were being given the labels that had been generated by the previous participant, and that those labels might be kind of strange if the previous person made errors or did something weird. The participants were asked to just do their best to learn the language, and to provide labels that would then be given to the next participant.

A critical element of this design is that the standard task was *exactly the same* across conditions. All of the data we analyze here is from that standard task (since the partner testing in the HIGH PRESSURE condition was only there to determine payment, and not relevant to our research question). The only difference between conditions that could affect the data we analyze is what participants thought their goal was and how they thought the stimuli were generated.

Results

The main question of interest is how much participants in each condition over-regularized by producing the **main** affix (or any single affix) more than 60% of the time. However,

[3]Stems were: dut, sil, zeg, mab, yok, pim, ren, jaf, wux, and cov and the affixes were: om, ep, ad, ig, and ut. Objects used were: babies, balls, beds, birds, books, cars, cats, cups, dogs, and shoes.

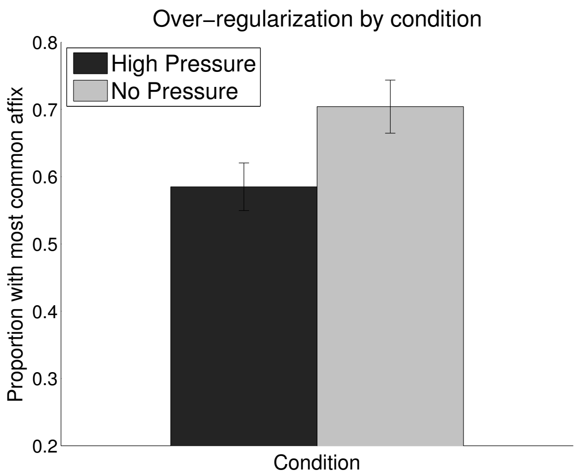

Figure 2: Proportion of time the most common affix is produced in each condition (after having occurred 60% of the time in the input). Participants in the NO PRESSURE condition over-regularized significantly more than did participants in the HIGH PRESSURE condition.

it is first necessary to determine whether any differences in over-regularization are associated with differences in overall performance or attention between conditions.

Accuracy. We can calculate an accuracy score for each condition, based on the percentage of stems that are generated in response to the appropriate item. A stem is counted as correct if it is no more than one letter different from the correct stem; thus, cav would be an acceptable variant of cov, but div would not be.[4] As Figure 1 demonstrates, people in both conditions were highly accurate, and there was also no significant difference between conditions ($t(47) = 0.807, p = 0.424$, two-tailed). This suggests that any differences in over-regularization did not result from people in the two groups paying different amounts of attention to learning the stems or affixes.

Over-regularization. The main question is if people show different levels of over-regularization in each of the two conditions. To determine this, we calculated the percentage of time that the most commonly-used affix was used by each participant. (For almost all participants, the most commonly-used affix was the **main** one; however, we did not want to presume that it always would be). As with the *accuracy* score, two affixes counted as the same if they differed by no more than one letter.[5]

Figure 2 shows the average percentage of trials in which participants produced the most common affix. It is evident that participants in the HIGH PRESSURE condition approximately probability match, while participants in the NO PRESSURE condition over-regularize significantly more ($t(47) = -2.250, p = 0.029$, two-tailed). This sort of population-level analysis can be misleading, however. It is possible that 60% of the people in the HIGH PRESSURE condition produce the most common affix 100% of the time and 40% of the people never produce any affix more than once. Individual performance can also be revealing about to what extent the conditions resulted in different strategies among the participants.

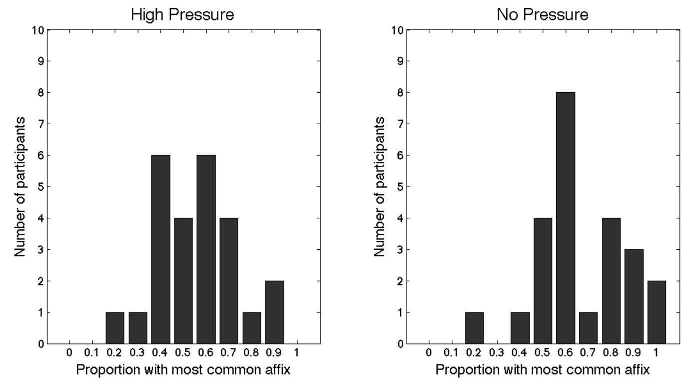

Figure 3: Histogram of the number of participants according to how often they produced the most common affix. It appears that in the HIGH PRESSURE condition, there is a unimodel distribution of responses centered near 60%. By contrast, the NO PRESSURE condition appears to have a bimodal distribution of one group that over-regularizes strongly and another group that probability matches.

Figure 3 therefore shows the histograms of each of the individual participants. Although such histograms are inevitably somewhat noisy and we should be careful about overinterpreting, they suggest two things. First, the number of participants who over-regularize in an extreme way (by producing the most common affix 80% of the time or more) is noticeably higher in the NO PRESSURE condition. In other words, these population differences actually do reflect the presence of more people who over-regularize strongly. The second interesting thing is that the distribution of participants in the HIGH PRESSURE condition is approximately unimodal: most people produce the most common affix between 40% and 70% of the time, and there are few outliers on either extreme. By contrast, the distribution in the NO PRESSURE condition looks more bimodal: there is one group of people who over-regularize strongly, and another set who do not. We will address some of the implications of this in the next section.

Discussion

Overall, this work indicates that adults over-regularize inconsistent linguistic input more often when placed in a situation in which the pressure for correct responding is reduced and they have reason to believe that the inconsistency does not correspond to an underlying regularity. Indeed, what strategy is sensible changes depending on whether one believes that the affixes covary according to some hidden regularity, or covary arbitrarily. If there is a regularity, it is sensible to try to find and match it; if not, it is sensible to remove the inconsistency. The NO PRESSURE condition, by emphasizing that the

[4] All analyses were also performed with a definition requiring the stem to match exactly; results were qualitatively identical.

[5] We also performed all possible combinations of two additional analyses. In the first, we used a more stringent definition of sameness, such that the affixes had to be identical to be counted. In the second, we considered only the subset of affixes for which the stem had been applied correctly. In all of these cases, participants in the NO PRESSURE condition over-regularized significantly more than participants in the HIGH PRESSURE condition.

input came from a previous participant who might have made errors, strongly implied that any variability was random or accidental. In contrast, the HIGH PRESSURE condition, by giving a cover story implying that the participant was a scientist whose job it was to learn the language, and especially by making payment contingent on successful "communication" of that language, created the strong implication that there was some regularity there to learn. It is interesting that increasing communicative pressure did not result in more removal of inconsistency in the language; again, this is probably because the participants thought that the inconsistency reflected an important regularity, even if they weren't sure what it was.

Our two conditions confound two effects: one is about the different assumptions participants might make about how likely the data is to contain errors, and the other is about how much pressure they feel to match some "correct" standard. We did not try to disentangle these things here because it would be very difficult, since simply knowing that the data might contain errors would decrease the pressure to correctly learn that data. The main point of this research is that differences in the impressions one brings to a task – whatever they may be – can change the degree of over-regularization.

Other than the experimental manipulations, the main implementational difference between this work and previous work is that both the input and the responses were written rather than auditory. However, since *both* conditions were written, the non-auditory nature of the experiment could not explain the difference between the conditions.

What does it mean that only some of the participants adopted an over-regularization strategy in the NO PRESSURE condition? One possibility is that those who did not adopt one still assumed that the inconsistencies may have been associated (perhaps more noisily) with regularities in the data. After all, it is a rather strange pattern of errors to have 10% each of four other kinds of affixes, but no other misspellings or incorrectly used stems. One way to investigate this possibility would be to design an experiment in which the affixes occur less often or there are additional errors in the input. However, since it is known that adults over-regularize more frequently when the inconsistency is less frequent (Hudson Kam & Newport, 2009) and any introduction of other errors would have to also occur in the HIGH PRESSURE condition in order to make the tasks equal, it is not clear that such a manipulation would be very illuminating. Still, this is a possibility we will consider for future work.

A potential limitation of this work is that the differences between the HIGH PRESSURE and NO PRESSURE conditions, although statistically significant no matter how the data is analyzed, are not large. In particular, even in the NO PRESSURE condition, not as many adult participants over-regularized as did children in Hudson Kam and Newport (2005) and Hudson Kam and Newport (2009). For this reason one of the first steps in future work will be to replicate these findings. That said, our results might underestimate the magnitude of the true effect to the extent that our participants didn't believe the "error" cover story but more of the children thought or assumed that the language they heard had some errors. Another quite likely possibility is that not all of the difference between children and adults is reducible to their different assumptions. Finally, any effect of naturally different assumptions between children and adults on the same task in Hudson Kam and Newport (2005) and Hudson Kam and Newport (2009) may have been exacerbated by subtle differences in their experimental design. In particular, while both adults and children learned a language called "Sillyspeak", only the children were told they were learning from someone who did not know the language themselves. It seems plausible that the children were far more likely to conclude that their input contained a lot more error, especially when combined with the very different social assumptions about the nature of laboratory studies that each group may have had.

What do these results mean for research showing that children over-regularize in non-linguistic situations as well as linguistic ones, as in Derks and Paclisanu (1967) and Jones and Liverant (1960)? Any application to non-linguistic domains must be made extremely cautiously, since many of the differences between the HIGH PRESSURE and NO PRESSURE conditions do not translate to a non-linguistic context. For instance, there is no obvious analogue of errors in a deck of cards or reinforcement pattern. In addition, these experiments do not have the learner/teacher dynamic that the linguistic ones do, which might be associated with very different patterns of assumptions. That said, children may still feel less pressure than adults to be "correct" or not look "stupid" to the scientists, so we cannot rule out the possibility that this plays a role even in non-linguistic domains.

Within the area of language, much of the interest in children's over-regularization arises because children and adults differ in their propensity to over-regularize outside of the lab as well as in it. Deaf children exposed to the inconsistent sign language of hearing parents will over-regularize that language and produce regular grammatical forms (Singleton & Newport, 2004), but 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). If over-regularization in children in these experiments is driven by differences in the assumptions they bring to the task, how do we explain these differences in real life?

In answer, we can only speculate. However, several possibilities present themselves. One is that children's over-regularization in real life is driven by some of the same factors that we are calling task demands here: that is, perhaps children just assume that more of their input is irregular or full of errors, or they are less bothered by trying to be "correct" or not look like an idiot and therefore focus on simply communicating clearly. Another possibility is that the different assumptions manipulated in the tasks here do not explain all of the differences between children and adults that are found in laboratory experiments. Still another possibility is that the kind of over-regularization measured in these experimental tasks does not map cleanly onto the over-regularization differences observed in natural language. At a minimum, one of these processes occurs in hours or days, and one takes years. Moreover, children may not bring the same assumptions to real language learning as to this kind of experimental task. Even beyond that, most variation in natural language is consistent in some way (e.g., Chambers, Trudgill, & Schilling-Estes, 2003). Thus, any child/adult differences in learning languages from native speakers[6] may not be traceable to the kinds of over-regularization differences found in these experiments. Thus, the implications for language acquisition must be speculative at this point, and we cannot say at this point for sure how to reconcile these findings with the language acquisition literature. It is important to note, however, that there are many ways they could be reconciled. Pursuing this is a project for future work.

This work may additionally have implications for adult second language acquisition, since it demonstrates that adults may change strategies in response to the nature of the explicit instructions they receive. This too is a project for future work.

The bottom line from this paper is that it is important to be cautious about drawing strong conclusions from existing laboratory studies to differences in how children and adults over-regularize when learning natural language. At the very least, the story is probably complex. For instance, some work shows that infants probability match in a looking time experiment (Davis, Newport, & Aslin, 2011), and even some of the original work in non-linguistic domains showed that children over-regularized the same amount as adults (Weir, 1964) or flipped strategies differentially depending on the nature of the reward (Stevenson & Hoving, 1964). The picture is therefore currently somewhat murky, even as regards the extent to which – and at what ages – children tend to over-regularize more than people of other ages.

In sum, this paper offers some reason to believe that over-regularization behavior can be driven by different assumptions about the goal of the experimental task and the origin of the data. If we are to understand what drives differences between adult and child language learning, we need to determine the extent to which the experimental findings in the literature stem from differences in such assumptions. It is also critical to further explore the extent to which these assumption explain differences in the learning of natural languages. This work is the first step along that path.

Acknowledgments

Thank you to Natalie May, Angela Vause, and Daniel Carabellese for recruiting participants and running the experiments. Thanks to them as well as Dan Navarro and Irina Baetu for useful discussions. This research was supported by ARC grants DE120102378 and DP110104949.

[6]Pidgins and creoles would be a different story, of course, because there the input might indeed be truly inconsistent. That said, there are many additional complexities in that case as well; for instance, adults receive input from different people than children, and have different goals in their communications in the kind of setting that causes pidgin languages to emerge.

References

Baum, W. (1975). Time allocation in human vigilance. *Journal of the Experimental Analysis of Behavior*, *23*, 45-53.

Baum, W. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior*, *32*, 269-281.

Bradshaw, C., Ruddle, H., & Szabadi, E. (1981). Studies of concurrent performances in humans. In C. Bradshaw, E. Szabadi, & C. Lowe (Eds.), *Quantification of steady-state operant behaviour.* Amsterdam: North Holland Biomedical Press.

Bradshaw, C., Szabadi, E., Bevan, P., & Ruddle, H. (1979). The effects of punishment on free-operant choice behavior in humans. *Journal of the Experimental Analysis of Behavior*, *31*, 71-81.

Castellan, N. J. (1974). The effect of different types of feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, *11*, 44-64.

Chambers, J., Trudgill, P., & Schilling-Estes, N. (2003). *The handbook of language variation and change*. Blackwell.

Davis, S., Newport, E., & Aslin, R. (2011). Probability-matching in 10-month-old infants. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

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

Herrnstein, R. (1961). Relative and absolute strength of responses as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267-272.

Herrnstein, R. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, *13*, 243-266.

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.

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.

Jones, M., & Liverant, S. (1960). Effects of age differences on choice behavior. *Child Development*, *31*, 673-680.

Perfors, A. (2011, .). Memory limitations alone do not lead to over-regularization: An experimental and computational investigation. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

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).

Pierce, W. D., & Epling, W. F. (1983). Choice, matching, and human behavior: A review of the literature. *The Behavior Analyst*, *6*, 57-76.

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

Ramscar, M., & Gitcho, N. (2007). Develomental change and the nature of learning in childhood. *Trends in Cognitive Sciences*, *11*(7), 274-279.

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.

Stevenson, H., & Hoving, K. (1964). Probability learning as a function of age and incentive. *Journal of Experimental Child Psychology*, *1*, 64-70.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*, 101-118.

Wearden, J. (1983). Undermatching and overmatching as deviations from the matching law. *Journal of the Experimental Analysis of Behavior*, *40*, 332-340.

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