THE ACQUISITION OF PRODUCTIVE RULES IN CHILD AND ADULT LANGUAGE LEARNERS

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Neuroscience

By

Kathryn Dolores Schuler, B.A.

Washington, DC
June 23, 2017
In natural language, evidence suggests that, while some rules are productive (regular), applying broadly to new words, others are restricted to a specific set of lexical items (irregular). Further, the literature suggests that children make a categorical distinction between regular and irregular rules, applying only regular rules productively during acquisition. This strong distinction has led to the central question explored in this dissertation: what governs the acquisition of productive rules in children? In the literature, a number of approaches have been proposed to account for the productivity of some rules, but most fail to capture this acquisition process adequately. This dissertation focuses on one model of productivity, the Tolerance Principle, which has been shown to accurately predict productive rule formation on a number of rigorous measures. The goal of this dissertation is to test the Tolerance Principle as a model of productive rule acquisition using artificial grammar learning experiments in children and adults. To this end, we conduct three experiments to assess whether the Tolerance Principle can predict productive rule formation in children and adults. Across these three experiments, we find that the behavior of children is well predicted by the Tolerance Principle model, but the behavior of adults is not. Thus, a secondary goal of the dissertation is to argue that the Tolerance Principle is a model of productive rule formation that is exclusive to children. We hypothesize that cognitive differences between children and adults, particularly memory and cognitive control differences,
may explain why the behavior of children but not adults is well predicted by the Tolerance Principle. We then demonstrate how these hypotheses can be tested in two further experiments with adults.

INDEX WORDS: Language Acquisition, Rules, Morphology, Syntax, Statistical Learning, Artificial Language Learning
DEDICATION

This thesis is dedicated to my daughter.
ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (HD37082 and DC014558) to Elissa Newport, by the Center for Brain Plasticity and Recovery at Georgetown University, and by the Feldstein Veron Innovation Fund.

I am grateful to everyone who has contributed to the work in this dissertation. Thanks to the Newport lab managers over the years, especially Katie Olson, Sarah Furlong, and Jaclyn Horowitz; to the rotation students and research assistants who helped with data collection and transcribing, especially John Davis, Breana Downey, Marie Beasley, Jessie Burke, and Rebecca Rennert; to the DC Metro area schools who allowed us to conduct research at their facilities, especially the Washington International School and the Arlington YMCA; and to the children and adults who participated in this research.

I am especially grateful to my advisor, Elissa Newport, who has made a huge impact on who I am as a scientist and the path my life has taken; to Charles Yang for his mentorship, for his perspective on language acquisition, and for the Tolerance Principle; and to my committee, Peter Turkeltaub, Michael Ullman, Ruth Kramer, and Barbara Landau for their helpful discussions and suggestions.

I am also grateful to the Interdisciplinary Program in Neuroscience for taking me in and teaching me Neuroscience; to the Cognitive Science group at Georgetown; and to the Language Acquisition (L1A) journal club, especially David Lightfoot, Ruther Kramer, Heidi Getz, Alix Handslshuh, Shannon Mooney, and Bryce Huebner for the amazing discussions and for shaping how I think about language acquisition.
Finally, I am forever grateful to my family and friends for their love and support, especially my parents and brother who taught me how to be a good person and how to work hard; and to my husband, who is a real-life super hero. I couldn’t have done this without you.
# Table of Contents

## Chapter

1 Introduction .................................................. 1  
   1.1 Productivity in natural language ......................... 1  
   1.2 The acquisition of productive rules ........................ 3  
   1.3 Overview of the dissertation ............................... 6  
   1.4 Common problems for models of productivity ............ 7  
   1.5 Previous approaches to productive rule formation ....... 11  

2 The Tolerance Principle (Yang, 2016) ............................ 20  
   2.1 Overview of the Tolerance Principle model .................. 20  
   2.2 Plausibility of model assumptions ........................... 21  
   2.3 Formalization of the model .................................. 26  
   2.4 Model performance on the common problems ............... 27  
   2.5 The need for converging evidence from child psycholinguistics 35  
   2.6 Overview of experiments ...................................... 36  

3 Testing the Tolerance Principle in children learning an artificial language 38  
   3.1 Introduction to Experiment 1 ................................. 38  
   3.2 Experiment 1 materials and methods ....................... 41  
   3.3 Experiment 1 results ......................................... 46  
   3.4 Experiment 1 discussion ...................................... 51  

4 Testing the Tolerance Principle in an artificial language with highly frequent exceptions 55  
   4.1 Introduction to Experiment 2 ................................ 55  
   4.2 Experiment 2 materials and methods ....................... 57  
   4.3 Experiment 2 results ......................................... 60  
   4.4 Calculating the personal Tolerance Principle: revised methods for Experiment 2 63  
   4.5 Experiment 2 revised results ................................. 65  
   4.6 Experiment 2 Discussion ...................................... 66  

5 Comparing evaluation metrics: the Tolerance Principle vs. the majority of types 69  
   5.1 Introduction to Experiment 3 ................................. 69
5.2 Experiment 3 materials and method ........................................... 71
5.3 Experiment 3 results and discussion ........................................... 75

6 Is the Tolerance Principle exclusive to children? ........................... 79
   6.1 Introduction ............................................................................. 79
   6.2 Maturational constraints on language acquisition ....................... 80
   6.3 The Tolerance Principle as a model of acquisition ....................... 83

7 Experimental approaches to adults and the Tolerance Principle .... 92
   7.1 Introduction ............................................................................. 92
   7.2 Adults acquiring a larger artificial language ............................... 93
   7.3 Adults performing the wug test under time pressure ................... 100
   7.4 Discussion ............................................................................... 106

8 Conclusion .................................................................................... 107
   8.1 Summary of the dissertation .................................................... 107
   8.2 Implications for acquisition of productive rules ....................... 108
   8.3 Implications for investigating maturational constraints in language acquisition ............................................... 109
   8.4 Implications for the study of language acquisition more broadly 110
   8.5 Future directions ................................................................. 111

APPENDIX
   A Experimental Stimuli ............................................................... 112
   B Full Derivation of the Tolerance Principle (Yang, 2016) ............... 116

BIBLIOGRAPHY ........................................................................... 118
List of Figures

1.1 Sample trial and results for the wug test adapted from Berko (1958) .......................... 5
3.1 Examples trials for exposure phase and production test in our experimental paradigm. .......................... 45
3.2 Usage of plural markers by children in Experiment 1 in the 5R4E and 3R6E conditions. .......................... 47
3.3 Usage of the regular form by individual children and adults in the 5R4E and 3R6E conditions in Experiment 1. .......................... 48
3.4 Usage of the regular form by children and adults compared with predictions made by the Tolerance Principle and the token frequency in Experiment 1. .......................... 50
4.1 Usage of plural markers by children in Experiment 2 in the 5R4E and 3R6E conditions. .......................... 61
4.2 Usage of the regular form (R) by individual children and adults in Experiment 2 in the 5R4E and 3R6E conditions. .......................... 62
4.3 Example trial for the revised rating test in Experiment 2. .......................... 64
4.4 Usage of the regular form (R) by individual children in Experiment 2 according to the prediction of their personal Tolerance Principle (Rule or No Rule). .......................... 67
5.1 Usage of plural markers by children and adults in Experiment 3, in which there are 10 regulars and 6 exceptions. .......................... 75
5.2 Usage of the regular form (R) by individual children and adults in Experiment 3, in which there are 10 regulars and 6 exceptions. .......................... 77
6.1 Gray matter volume for structures underlying the declarative and procedural memory circuits across development (ages 7 to 24 years), adapted from Wierenga et al. (2014) .......................... 88
7.1 Usage of the regular form by adults in each version of Experiment 4. .......................... 98
7.2 Example trial for the timed production test in Experiment 5. .......................... 102
7.3 Usage of the regular form by individual adults in the 5R4E and 3R6E conditions in Experiment 5. .......................... 105
List of Tables

2.1 Tolerance Principle threshold of productivity for the English past tense as vocabulary size increases over time. .............................. 29
3.1 Frequency of nouns and their plural markers in Experiment 1, Language A. ................................................................. 43
4.1 Frequency of nouns and their plural markers in Experiment 2, Languages A and B. .......................................................... 59
4.2 Calculation and prediction of each child’s personal Tolerance Principle based on the number of noun-marker pairings the child has learned and the number of those that are exceptions. ......................... 66
5.1 Threshold of productivity the Tolerance Principle and Majority of Forms (MoF) evaluation metrics. ................................. 70
5.2 The number of regular forms required for productive rule formation for the Tolerance Principle and the Majority of Forms evaluation metrics. ...................................................... 72
5.3 Frequency of nouns and their plural markers in Experiment 3, Language A. ................................................................. 74
6.1 Tolerance Principle threshold of productivity for increasing values of \(N\) from Yang (2016). ......................................................... 84
7.1 Frequency of nouns and their plural markers in Experiment 4 for the High-Frequency Regular (HFR), Matched Token Frequency (MTF), and Farther From Threshold (FFT) conditions. ......................... 95
7.2 Token frequencies of the regular form in Experiment 4. .............. 97
7.3 Summary of t-test results in Experiment 4. ............................... 99
7.4 Summary of t-test results in Experiment 5. ............................... 104
A.1 Experimental stimuli for the 9 and 16 noun artificial languages. ... 113
A.2 Experimental stimuli for the 36 noun artificial languages ............ 114
A.3 Experimental stimuli for the production tests. .......................... 115
Chapter 1

Introduction

1.1 Productivity in Natural Language

When children learn a language, they do not simply memorize all of the words or sentences they are exposed to. Instead, they acquire the patterns by which these words and sentences are formed. We call these patterns rules. For example, in English, we apply the rule "add -d" to make the past tense of verbs (e.g., *walked*, *jumped*, *arrived*, *played*) and the rule "add -s" to make nouns plural (e.g., *apples*, *dogs*, *friends*). Further, these regular patterns are known as productive rules because they apply productively to new or novel words. For example, when newly minted verbs like *google* or *instagram* enter our vocabulary, productive rules allow us to immediately generate their past tense forms: *googled* and *instagrammed*. Similarly, if asked to provide the plural or past tense of a made-up word like *wug* or *rick*, even very young children are able to apply the regular pattern to generate *wugs* and *ricked* (Berko, 1958). While there are a number of exceptions to the "add -d" and "add -s" rules – the past tense of *go* is *went*; the plural of *goose* is *geese* – the overwhelming majority of English words obey these regular, productive patterns.

There is substantial evidence to suggest that these regular patterns are indeed productive. Following Berko (1958) mentioned above, a number of studies have found that both children and adults generalize the regular form to novel stems in experiments (Albright & Hayes, 2003; Ambridge, 2010; Bybee & Moder, 1983; Gordon &
Miozzo, 2008; Prasada & Pinker, 1993; Ramscar, 2002). Further, when adults occasionally make language errors, these errors are often an over-application of the regular form (Stemberger, 1983), a phenomenon know as overregularization (e.g., accidentally producing "breaked" as the past tense of "broke"). Verbs derived from nouns – as is the case for google and instagram mentioned above – are inflected with the regular form, even if the stem from which the word is derived is itself irregular (e.g., the past tense of fly out is flied out) (Huang & Pinker, 2010; Kim et al., 1991, 1994). Finally, throughout history, thousands of verbs that were once irregular have, over the course of language evolution and change, become regularized (Lieberman et al., 2007).

Importantly, however, not all linguistic rules are productive. Some inflected forms are idiosyncratic to a single lexical item, as in the go/went example above. In addition, some rules apply to only a restricted subset of lexical items, like sing/sang/sung and ring/rang/rung, but are not productive in that they do not apply broadly to new words. Rating and production test studies show that, for novel items, adults will generalize by analogy to an irregular form only under very limited circumstances (Albright & Hayes, 2003; Ambridge, 2010; Bybee & Moder, 1983; Bybee & Slobin, 1982; Prasada & Pinker, 1993), and judgment studies reveal that the regular form is also acceptable in these cases. Moreover, when adults do make such overirregularizations in production studies, these errors have been argued to be merely a result of the task and not necessarily a reflection of productivity in natural language\(^1\) (Schütze, 2005, e.g.). When adults make unexpected overgeneralization errors, they only very rarely overirregularize, making overregularization errors much more often (Bybee,

---

\(^1\)In the literature, the sneak-snuck overregularization is often used to claim that analogical errors are common (Bybee & Moder, 1983). However, sneak/snuck appears to be an isolated case in which an irregular pattern in extended productively. While both sneak/ and snuck are currently considered acceptable past-tense forms of sneak (Anderwald, 2013), it has been hypothesized that snuck has historically been present in the input (Yang, 2016), and is thus not an overgeneralization of the irregular form.
1985; Pinker, 1995). This pattern of overgeneralization is also the case in second language acquisition; overregularization is much more common than overirregularization (Pica, 1983). Finally, historically only a handful of forms that were once regular have become irregular (e.g., wear, spit, dig), a far cry from the thousands of forms that have been regularized throughout history (Bybee, 1985).

1.2 The acquisition of productive rules

Taken together, the results from the previous section suggest that natural languages make a strong distinction between rules that are productive and rules that are not. Importantly, the distinction between productive and unproductive rules is even more pronounced in the acquisition literature. Children have long been observed to make overgeneralization errors in their own productions as they acquire language (e.g., "Daddy goed to the store.") (Chamberlain, 1906; Ervin, 1964; Ervin & Miller, 1963). While the rate at which children make these overregularization errors was once thought to be quite low – 2.5% of children’s productions in Marcus et al. (1992) and 4.2% in Pinker (1995) – recent studies have found this rate to be much higher. Maslen et al. (2004) found 7.8% of children’s productions to be overgeneralizations of the regular form, while others have found the overregularization rate to be as high as 10% (Maratsos, 2000; Yang, 2002; Hoeffner, 1997). In contrast to this relatively high rate of overregularization, the rate at which children overgeneralize irregular forms turns out to be quite low. While analogical errors like bite/bote and wipe/wope are frequently cited examples of overgeneralization in the acquisition literature (Ambridge et al., 2015; Bowerman, 1982; Bybee, 1985; Pinker, 1999; Pinker & Prince, 1988), leading to the impression that such errors are quite common, the empirical data on rates of overirregularization suggest children almost never make mistakes of this kind.
For instance, Xu & Pinker (1995) found that children overgeneralize the irregular form on only 0.02% of productions\(^2\). Yang (2016) found a similarly low rate of overirregularization. After thoroughly analyzing CHILDES (MacWhinney, 2000), a corpus of nearly 2 million tokens of child speech, Yang (2016) found zero examples of overgeneralizations of the irregular form. Moreover, evidence that children somewhat frequently overgeneralize the regular form while almost never overgeneralizing the irregular forms has been observed across many languages. Studies investigating the acquisition of the German participle system have found that 10% of children’s productions are overgeneralizations of the regular form while only 0.75% are overirregularizations (Clahsen, 1999; Clahsen & Rothweiler, 1993; Weyerts & Clahsen, 1994). Similarly, in a study of the acquisition of Spanish verbal inflection, the rate at which children overgeneralized the irregular form was just 0.001% (Clahsen et al., 2002; Mayol, 2007). Perhaps the only overirregularization error that occurs with any frequency during language acquisition is "brang". However, this particular error has been attributed to the fact that "brang" is thought to be attested in the input, either due to variations in dialect (Herman & Herman, 2014) or simply as a part of child-directed speech (MacWhinney, 2000).

Further, this pattern is observed in experimental investigations as well. While children consistently apply the regular form productively to novel items in experiments (Albright & Hayes, 2003; Ambridge, 2010; Berko, 1958; Bybee & Moder, 1983; Gordon & Miozzo, 2008; Prasada & Pinker, 1993; Ramscar, 2002), children almost never apply irregular forms in an analogous way. For example, when prompted to pro-

---

\(^2\)Xu & Pinker (1995) use the term "weird errors" for this 0.02% to indicate productions that could not be classified as either correct or an overgeneralization of the regular form. While this may include possible overgeneralizations of irregular forms, Yang (2016) suggests that many of these "weird errors" may simply be speech errors (e.g. the child says "fit" when she means to say "feet") or due to other processes like -t/d deletion (Labov, 1972, 1989), in which the child produces "slep" (not "slept") as the past tense of "sleep".
Figure 1.1: **Sample trial and results for the wug test adapted from Berko (1958).** Sample trial to elicit the past tense of "gling" is shown in (a) and percentage of children providing the inflected form for wug, bik’s, zibbing, ricked, and glang are shown in (b). Note that children apply the regular form consistently for plural, possessive, progressive, and past tense, but do not apply the analogous irregular form "glang" when prompted for the past tense of "gling".
vide the past tense of "gling", analogous to ring/rang and sing/sang, only 1 child out of 86 produced "glang", an irregular form analogous to "ring/rang" or "sing/sang" (Berko, 1958) (see Figure 1.1).

1.3 Overview of the dissertation

The literature reviewed above provides strong evidence that there is a categorical distinction between rules that are productive (regular) and rules that are not (irregular). Further, that children are highly sensitive to this distinction during language acquisition, applying only the regular, and not the irregular, rules productively. The tendency to generalize some rules but to restrict others motivates the question: what governs the formation of productive rules during language acquisition? That is, how do children determine which of the many rules they are exposed to can be applied productively?

In this dissertation, I provide evidence from artificial language learning experiments that one particular model, the Tolerance Principle (Yang, 2005, 2016), accurately predicts productive rule formation in acquisition. With the remainder of this chapter, I review some of the most common problems discussed in the literature on the acquisition of rules and productivity, and the models that have been proposed to address them. In Chapter 2, I introduce the Tolerance Principle, a recent model of productive rule formation proposed by Yang (2016), and review the evidence for its efficacy as a model of productivity. In Chapters 3–5, I test the Tolerance Principle across three artificial language learning experiments conducted with both children and adults. In Chapter 6, I discuss the differences between adults and children in our experiments and what this means for the Tolerance Principle as a model of acquisition. In Chapter 7, I conduct two further experiments with adults to explore the
difference between adults and children proposed in Chapter 6. Finally, in Chapter 8, I will discuss the conclusions that we can draw from the experiments presented in chapters 3, 4, 5, and 7, as well as the broader impacts of both the Tolerance Principle as a model of productive rule formation and conducting artificial language learning experiments with children.

1.4 Common problems for models of productivity

Before we review the previous approaches to productivity in acquisition, it is helpful to review some of the common problems these approaches need to solve in order to be considered adequate models of productive rule formation. In the following sections, we will consider in turn: The U-shaped development of English past-tense, error patterns and overregularization in language acquisition, and the case of German plural.

1.4.1 The U-shaped development of English past-tense

In language acquisition, researchers have long observed that, before children begin to make overgeneralization errors, they often experience an initial "correct period", during which they provide the correct irregular form (e.g. "daddy went to the store"). This period is followed by a period of overgeneralization (e.g. "daddy goed to the store"), after which children ultimately arrive on the correct irregular form again (Cazden, 1968; Kuczaj, 1977; Marcus et al., 1992). Though this U-shaped pattern has been historically contested, particularly by proponents of the connectionist approach (e.g. Hoeffner, 1997; MacWhinney & Leinbach, 1991; Plunkett & Juola, 1999; Plunkett & Marchman, 1993, 1996), it is generally agreed that the U-shaped trajectory exists for at least some verbs in at least some children. Thus, one way to assess the efficacy of a model of rule formation is to determine whether the model can generate this same
U-shaped pattern of results. As we will see in the following section, the connectionist models have been particularly thoroughly tested on this problem and fail to yield results demonstrating a similarly U-shaped trajectory. Only a small number of these models have revealed this U-shaped pattern, and in those cases, the models were provided with learning circumstances that would be unrealistic to assume for a child (Hoeffner, 1997; Plunkett & Marchman, 1993).

1.4.2 Error patterns and overregularization in language acquisition

In the previous section, we saw that children treat productive rules categorically, evidenced in part by the fact that they frequently overgeneralize the regular form, but almost never do so with irregular forms. Empirical evidence suggests that the rate of overregularization is around 10%, while the rate of overirregularization is 0.02% or less. As in the previous section, one way to assess the validity of models of productivity is to determine whether these models can produce similar error patterns to children. As we will see in the following section, while some models are able to match this rate of overregularization quite well, these models also produce overirregularization rates that are significantly higher than those observed in children (Marcus, 1995; Johnson et al., 2007; O’Donnell, 2011a).

Further, while we have seen that error patterns for regular rules are somewhat high (10%), and those of irregular rules are extremely low, it turns out that the most common type of errors children make with inflectional morphology are errors of omission. That is, when children make mistakes during language acquisition, they most often provide no form at all (see Yang, 2016, for review). Further, this error pattern is consistent across languages. For example, in a study investigating the acquisition of German agreement affixes, the child under observation makes an error of omission – that is, omits an affix under circumstances in which supplying an affix is obligatory –
on 17% of her productions. When the child does supply an affix, she does so correctly 98% of the time (Clahsen & Penke, 1992). A similar pattern of omission is observed for Italian (Caprin & Guasti, 2009; Guasti, 1993; Pizzuto & Caselli, 1994) and languages with more complex morphology, such as Swahili (Deen, 2005), Inuktitut (Allen, 1996), and Bantu languages (Demuth & Nurse, 2003). In a study of agreement in Xhosa, 97.2% of all errors children make are those of omission (Gxilishe et al., 2007). This pattern of results suggests that, when children have neither formed a productive rule nor acquired or memorized the irregular form, they provide no inflection, preferring simply to produce the bare form of the stem.

Yang (2016) points out that this behavior is reminiscent of the paradigmatic gaps that can occur in natural language, in which there is no productive or default form (Halle, 1973). A frequently cited example is the past tense in Russian, in which there are nearly 100 verbs that "lack first person singular forms of the past tense" (Halle, 1973, page 7). For example, the sentences in 1 (from Halle, 1973) are considered ungrammatical by native Russian speakers, and are thus not attested in language corpora, even though there is no good explanation (syntactic, semantic, phonological, or otherwise) as to why this should be the case.

(1) a. *lažu "I climb"
   b. *pobežu "I conquer"
   c. *deržu "I talk rudely"
   d. *erunžu "I behave foolishly"

It turns out that circumstances in which there is no productive form are relatively common in the world’s languages (see Baerman et al., 2010; Fanselow & Féry, 2002; Rice & Blaho, 2009, for reviews). In English, for example, there is no past participle form for *stride. While native speakers judge strode to be its past tense form, they
are reluctant to accept *stridden, strode*, or *strided* as the past participle (Pullum & Wilson, 1977; Pinker, 1999), suggesting this is an arbitrary gap in the lexicon. What is more, children exposed to languages in which these paradigmatic gaps exist correctly acquire these arbitrary missing forms. For example, in Polish, masculine nouns in the genitive singular can take either the -a or -u inflection, but neither is the productive (or default) form. Accordingly, children acquiring Polish do not make overgeneralization errors with either the -a or the -u forms in their productions (Dabrowska, 2001).

Thus, models of productive rule acquisition need to reproduce this pattern of errors as well. Like children producing novel forms, models should get "stuck" in a paradigmatic gap when they have not acquired enough evidence to form a productive rule, producing the bare stem (or a null affix) instead of guessing at the default form. As we will see in the following section, reproducing this behavior is especially problematic for many models, in which the very nature of the approach is to "pick a winner" between the regular and irregular forms. While some of these competition based models do suggest that, when neither the regular nor an irregular form has been acquired the learner will be forced to provide the bare stem (e.g. Pinker, 1999), many others provide no such avenue by which an error of omission could arise.

### 1.4.3 The case of German plural

An often cited observation in the linguistics literature is that forms that are not productive, like the irregular forms and exceptions, tend to have lower type frequency but higher token frequency, while forms that are productive, like the regular forms, tend to have the opposite pattern (Baayen, 1992, 1993; Baayen & Lieber, 1991; Baayen & Renouf, 1996; Barnwell, 2010; Bauer, 2001, for example). This observation has lead to the hypothesis that perhaps the productive rule is determined by the pattern that
applies to the majority of types (e.g. Bybee, 1995). While this happens to be true for the English past tense, researchers have long pointed out that this approach is insufficient to capture productivity cross-linguistically. Indeed, there are a number of examples in natural language that violate this pattern. In particular, there are several languages in which a rule is productive even though it does not apply to the majority of types. Among these, the most widely cited example is the case of German plural, in which the productive form, -s, is actually the least frequent allomorph in both type and token frequency (Janda, 1990; Marcus et al., 1995). Arabic sound plural (Forrester & Plunkett, 1994; Plunkett & Nakisa, 1997) and Polish inflectional marking (Dabrowska, 2004) have similarly low-frequency productive forms. Furthermore, the opposite is also attested in natural language. That is, a rule can fail to be productive even it has a high type frequency. For example, in Dutch -ster is a suffix added to make nouns feminine while ver- is a prefix added to construct verbs. Interestingly, the -ster suffix is productive while the ver- prefix is not, even though type frequency of ver- is much higher than the type frequency of -ster (Baayen, 2009).

1.5 Previous approaches to productive rule formation

A number of approaches to rule learning and productivity have previously been proposed in the literature. In the following section, we review these existing accounts, comparing the predictions of these models to the common problems discussed in the previous section. We consider in turn, the connectionist approaches, rule-based accounts, and quantitative approaches to productivity.
1.5.1 Connectionist approaches

In contrast to the idea that children must acquire an abstract rule in order to correctly inflect the past tense, Rumelhart & McClelland (1986) proposed a pattern-association model, also known as a connectionist model, that took an alternative approach. Using the English past tense as a test case, Rumelhart & McClelland (1986) proposed that children could learn to correctly inflect verbs via a simple association mechanism, by which features of the stem are learned to be associated with features of the past tense form. For example, a connectionist model is trained, using a large number of past tense pairs, to associate phonological features of the stem with phonological features of the past tense form. As a performance test, the model is given a bare stem it did not experience during training and must output its past tense form.

Connectionist models such as Rumelhart & McClelland (1986) have been met with substantial criticism, particularly from proponents of rule-based approaches to language acquisition. Among these criticisms is the fact that these models fail to produce the error patterns characteristic of child language acquisition, particularly the u-shaped development of English past-tense (see Section 1.4.1). While these models have been shown to overregularize at approximately the same rate as children (10%), rates of overirregularization are significantly higher than those observed in children. Further, these models fail to exhibit the U-shaped trajectory of past tense development. Since Rumelhart & McClelland (1986), a number of researchers have attempted to produce the U-shaped function in connectionist models, by making changes or enhancements to the pattern-association networks (Hoefnner, 1997; MacWhinney & Leinbach, 1991; Plunkett & Juola, 1999; Plunkett & Marchman, 1993, 1996, 1991), but these attempts have largely failed. Only when provided with unrealistic input conditions, for example, receiving only irregulars at the beginning of training, can
some models reproduce this pattern for a small number of verbs (Hoeffner, 1997; Plunkett & Marchman, 1993).

1.5.2 Rule-based accounts

At the opposite extreme to the no-rule pattern-association models are the classic deterministic theories in which everything, including the irregulars, are abstract rules (e.g. Chomsky & Halle, 1968; Kiparsky, 1982; Halle & Marantz, 1993; Mohanan, 1986). These approaches are unsatisfactory for different reasons. First, there is fairly strong evidence to suggest that, while the regular form is generated by an abstract rule, the irregular forms and exceptions are stored in memory (e.g. Beck, 1997; Prasada et al., 1990; Seidenberg & Bruck, 1990; Ullman, 1999, 2001a). This work is reviewed in more detail in Chapter 6. Further, even if one is unconvinced by these arguments and prefers to believe that everything is generated by rules, in many cases these theories propose that more than one rule may apply to a given circumstance, and therefore the learner requires a mechanism by which to select the optimal rule.

To be fair, many of the deterministic theories do indeed attempt to explain how this competition between rules is resolved. For example, several approaches including those proposed in distributed morphology, rely on Panini’s principle, also known as the Elsewhere Condition. This principle states that a more specific rule will be preferred over a rule that is more general (Anderson, 1969; Brown & Hippisley, 2012; Halle & Marantz, 1993; Kiparsky, 1973; Stump, 2001). Thus, when an irregular rule is available it will be preferred over the more general regular rule. In another approach, the lexicon is organized into levels, or strata, and rules that apply on earlier levels take precedence over rules that apply on later levels. Under this account, the irregulars are thought to be applied at an earlier level and, as such, the regular rule is only applied
if no irregular rule has been implemented on an earlier level (Halle & Mohanan, 1985; Kiparsky, 1982; Mohanan, 1986).

Another class of rule-based theories are known as the dual-route approaches, in which the regulars are generated via an abstract rule, as in the classic deterministic approaches above, but the irregulars are stored in memory (e.g. Clahsen, 1999; Pinker, 1999). As mentioned above, there is compelling evidence in support of this hypothesis from the psycholinguistic literature (Beck, 1997; Prasada et al., 1990; Seidenberg & Bruck, 1990; Ullman, 1999, 2001a). Still, this approach suffers from the same basic problem as the previous deterministic approaches - there is no mechanism by which to resolve conflict when both the regular and the irregular forms are available. To address this, many proponents of the dual route approach adopt a variant of the Blocking-and-Retrieval Failure hypothesis (Marcus et al., 1992). This hypothesis states that when both the regular rule and an irregular form are available, the irregular should apply (blocking principle)\(^3\). However, if the learner fails to retrieve an irregular form (retrieval failure), then there is nothing to block, and the regular form will therefore be applied. In this latter case, if the stem the learner is attempting to inflect is an irregular, an overregularization error will occur. However, if the learner has acquired neither the regular rule nor the appropriate irregular form, she will simply produce a bare stem, resulting in an error of omission (Pinker, 1999).

While the rule-based approaches appear to offer a reasonable framework from which to think about regulars, irregulars, and competition between them, a number of issues remain unresolved. First, how does one distinguish between the various rule-based theories? As O’Donnell (2011a) points out, the mechanisms for resolving competition among these theories are quite similar, and none provide a precise quantitative

\(^3\)Note that the Blocking-and-Retrieval Failure hypothesis is quite similar to the Elsewhere Principle
prediction about the learning trajectory of these processes with which to generate empirical comparisons. To address this problem, a number of models have attempted to, and been reasonably successful at, quantifying the degree of competition between competing forms in rule-based approaches (Taatgen & Anderson, 2002; Anderson & Lebiere, 1998; Yang, 2002; Albright & Hayes, 2003). Still, a second more glaring issue remains unsolved: we still have no mechanism with which to govern productive rule formation. The rule based approaches are concerned with whether children should apply the regular rule or an irregular form for any given stem, but precisely how children form the productive rule or decide that a particular form is indeed the regular form is simply not addressed.

Given the historical context in which the rule-based approaches were generated, it is not altogether surprising that the productive status of rules is not directly considered. The original question these approaches were designed to address was whether abstract rules were a necessary component of inflectional morphology and, more broadly, human cognition in general (Pinker, 1999; Marcus, 2001; Ullman, 2001a; Lima et al., 1994; Nooteboom, 2002; McClelland & Patterson, 2002). Indeed, the "past tense debate" is often discussed as a simple and useful test case for the necessity of rules. However, how these rules are formed and how children determine whether a given rule is productive - the central question of this dissertation - was not necessarily the primary concern of these early approaches.

4 There are a number of additional models based on inducing rules from phonological forms (Yip & Sussman, 1996, 1997; Molnar, 2001; Ling & Marinov, 1993; Mooney & Calif, 1995), but these are also merely finding potential rules and not necessarily determining the productivity of such rules.
1.5.3 Quantitative approaches to productivity

On the other hand, a number of linguists have attempted to directly address this question by precisely quantifying the productive process, that is, providing a way in which to compute the productive status of a rule. Among the simpler of these quantitative metrics, Bybee (1995) computes a productivity index based on the proposal that productivity is predicted by the frequency of an affix. However, as we saw in Section 1.4.3, this approach fails to capture productivity cross-linguistically, as there are a number of languages in which a morphological process is productive even though it has very low type frequency (e.g. German plural affix -s: Janda, 1990; Marcus et al., 1995), while others with high type frequency fail to become productive (e.g. Dutch verbal prefix -ver: Baayen, 2009).

However, other early approaches have used more complex metrics of productivity not necessarily based on the majority of types. Among these, Aronoff (1976) proposed that the productivity of a given form – the word-form rule (WFR) as he refers to it – can be quantified as the proportion of existing word types which take that form divided by the number of word types that could potentially take that form. Baayen & Lieber (1991) later formalized this as the equation in 1.1, where I is the index of productivity, V is the number of actual types taking a given form (WFR), and S is the number of types the WFR could potentially give rise to.

$$I = \frac{V}{S}$$ (1.1)

The problem here, pointed out by both Aronoff (1976) and Baayen & Lieber (1991), is twofold. First, one can not know the value of V with any certainty. It can only be estimated from some fixed corpora, and this estimate will likely miss some productive processes due to sampling effects. Second, due to the recursive nature of
natural languages, the number of potential words types is, in principle, infinite. Thus, regardless of the value of \( V \), the index of productivity, \( I \), will always also equal zero because it is computed via a division by infinity.

Following this, Baayen (1989) suggested a revised measure of productivity which makes use of the fact that the token frequency of a form is correlated with productivity (see also Baayen & Lieber, 1991; Baayen, 1992). In the following equation, \( p \) is the index of productivity, \( n_1 \) is the number of words formed by the process occurring in a corpus precisely once (a phenomenon known as the hapax legomena), and \( N \) is the total token frequency of words created by the morphological process in the corpus.

\[
p = \frac{n_1}{N}
\]  

(1.2)

This approach is problematic for many reasons (see Bauer, 2001, for review). Primarily, as with type frequency, the token frequency of the regular form is not always correlated with productivity (Janda, 1990; Marcus et al., 1995, e.g. German plural -s: ). Baayen (1993) has since revised his approach to incorporate both type and token frequency into the index of productivity. However, these revisions have continued to depend heavily on the hapax legomena phenomenon, and it is not clear that the hapax legomena is at all correlated with productivity. Though Baayen & Lieber (1991) argue that the hapax legomena captures the approximate productivity of a process – because the more word types a process applies to, the more likely that a few of those types have only been sampled once in a given corpus – many have expressed doubt that the hapax legomena is necessarily relevant to productivity (see Bauer, 2001). Bauer (2001) points out that the most obvious of these doubts was well summarized by Van Marle (1992, page 156) in a commentary on Baayen’s formula:
"I do not see what kind of direct relationship there is between the chance that a rule is put into action and the frequency with which the words that have already been produced by that rule are used. Once a word is coined, the frequency of the use of that word, it seems to me, is more or less irrelevant to the degree of productivity of that rule."

Regardless, each of these models suffers from larger problems. In particular, they have not been tested against a range of productive processes in a range of languages, nor have their predictions been compared to the acquisition data reviewed above. There are a few additional probabilistic models of productivity that have undergone these rigorous tests (e.g., O'Donnell, 2011a,b), but these models appear to suffer from the same problems as the connectionist models. While they are able to closely match the overregularization rate of children acquiring language, they overgeneralize the irregular form substantially more than children do.

Further, and perhaps more importantly, the computations that underlie each of these quantitative approaches result in an index of productivity that lands somewhere along a continuum between the values of 0 and 1, with 0 indicating that a rule is not at all productive and 1 indicating that it is completely productive. These gradient values imply that rules can have some degree of productivity, and, as such, fail to capture the categorical nature of productive rule formation, a view that is strongly supported by the acquisition literature reviewed throughout this chapter.

Finally, as Yang (2016) points out, these approaches are, at best, statistical models of the data. While researchers may be capable of using approaches such as these on large corpora to estimate the productivity of morphological processes in natural language, it is unlikely that young children have access to the complex variables that underlie these computations (e.g. the hapax legomena). And even if one could
reasonably assume that they do, it is not clear how a child would apply the resulting estimate of productivity to decide whether or not a rule should be productive. What is the threshold index of productivity for which children can assume a rule applies to novel situations? An index of 0.50? An index of 0.75?

Thus, what is missing from the previous literature is a model that captures how children determine, based on the language input they have received, whether or not a rule is productive. Such a model should be simple enough that we can assume even very young children have the cognitive capacities to engage in it. Further, when evaluated against the common problems reviewed in Section 1.4, the model should perform in a way that is consistent with children acquiring natural language. In the following chapter, I introduce one such model, the Tolerance Principle, which satisfies the conditions we have outlined above.
2.1 OVERVIEW OF THE TOLERANCE PRINCIPLE MODEL

The Tolerance Principle (Yang, 2005, 2016) is a learning model that quantifies the precise conditions for productive rule formation during language acquisition. It hypothesizes that a general rule will be formed when doing so is computationally more efficient than storing lexical forms individually. The model computes this computational efficiency by calculating the time complexity required for forming a rule with the time complexity required for accessing individual lexical forms. To illustrate, imagine that a learner is faced with a potential rule – for example, the English rule "add -d" to make the past tense of verbs. The English learner has encountered many items that obey this rule (the regular forms) as well as many items that do not (the irregular forms or exceptions). The Tolerance Principle assumes that, to be maximally efficient in forming the past tense of verbs, the learner can do one of two things:

(2) **Store all lexical forms individually**: store every item individually in a frequency-ranked list, searching the list every time there is an occasion to express the past tense of a verb.

(3) **Form a productive rule**: store only the exceptions in a frequency-ranked list. To express the past tense, the learner searches the list of exceptions first. If the target verb is not among these exceptions, the learner applies the rule.
The Tolerance Principle computes the time complexity required for each of these operations and assumes that the learner will adopt the optimal (i.e., faster) strategy. Productive rules, then, are formed only when it is more computationally efficient for the learner. Importantly, the decision to form a productive rule under the Tolerance Principle is a categorical one. Learners either form a productive rule or do not, depending on whether forming a productive rule is more computationally efficient than not forming one.

2.2 Plausibility of model assumptions

To compute the time complexity required for each of these operations, the Tolerance Principle makes a number of assumptions regarding the storage and retrieval of morphological forms as well as particular biases the learner may have that are relevant for the acquisition of productive rules. Below, we review these assumptions and address whether or not they are supported by sufficient empirical evidence.

2.2.1 The Elsewhere Condition

The Tolerance Principle makes use of the Elsewhere Condition, a traditional approach in linguistics for handling regular and irregular forms (Anderson, 1969; Brown & Hippisley, 2012; Halle & Marantz, 1993; Kiparsky, 1973; Stump, 2001). Also reviewed in Section 1.5.2, the Elsewhere Condition states that when a more specific rule (or form) is available, it is preferred over a more general one. For example, according to the Elsewhere Condition, an irregular form such as went will be preferred over the most general rule "add -d" when forming the past tense of go. The Tolerance Principle implements the Elsewhere Condition computationally as a serial search procedure, such that all of the irregular, or more specific, forms must be traversed
before the general rule is applied. This implementation can be thought of as a series of conditional statements: If a given stem matches *go*, then inflect as *went*; if it matches *sing*, inflect as *sang*; if it matches *fly* inflect as *flew*; otherwise (or elsewhere), apply the general rule "add -d".

Further, the Elsewhere Condition is implemented such that the "list" of exceptions is searched in serial order, according to the frequency rank of the word. While this is a reasonable theoretical assumption, is it plausible to assume that human learners engage in this kind of serial search procedure? It turns out that there are a number of empirical reasons to support this assumption. First, there are well known frequency effects for processing irregular forms, in which high frequency irregulars are processed significantly faster than low frequency irregulars (Beck, 1997; Prasada et al., 1990; Seidenberg & Bruck, 1990; Ullman, 1999, 2001a). Further, evidence from English past tense suggests that, for irregular forms, lexical decision time is significantly correlated with frequency rank (Balota et al., 2007; Yang, 2016).

Still, it seems quite computationally demanding for learners to maintain a list of all of the irregular forms and their frequency in the lexicon. Is it reasonable to assume the learners, especially young children, are engaging in such a demanding process? It turns out that this is not as demanding as one might suspect. Yang (2016) points out that "such a list needn’t be constructed by keeping track of all frequencies and then rearranging them in descending order." Instead, the learner simply requires a method of approximating the frequency rank of the irregulars. For example, Yang (2016) suggests a simple "move to front" process in which a word moves to the top rank position each time it is used, or a "move up" process (Rivest, 1976) in which a word is swapped in rank with the one just above it each time it is used. Either process would be simple to maintain computationally, and both are thought to closely approximate
relative frequency (Sleator & Tarjan, 1985). Yang (2016) illustrates the effectiveness of such approaches with the following example:

"Readers only need to look at their smartphones to be convinced. Double tapping the home button on an iPhone will reveal the list of active apps sorted by recency – which closely matches their relative frequency of usage."

A second assumption implied by the Tolerance Principle’s implementation of the Elsewhere Condition is that the learner must first search through their list of irregulars before applying the general rule. That is, the regulars have to wait for the exceptions. Yang (2016) himself points out that "this seems absurd." It certainly does not seem efficient to have to run through a long list of irregulars before a rule can be applied. Surprisingly, however, there is at least some evidence to support this assumption. In a German production study of the past participle, both children and adults produced the irregular suffix, -a, significantly faster than the regular form (Fleischhauer & Clahsen, 2012; Clahsen et al., 1997, 2004; Yang, 2016). Similarly, in a lexical decision task of German plural, lexical decision times for the irregular suffix, -er, were significantly faster than for the regular suffix, -s (Sonnenstuhl & Huth, 2002). Beyond the domain of morphology, a similar "irregular advantage" has been observed for idioms over literal meanings or compositional phrases (e.g., kick the bucket vs. lift the bucket) (Bobrow & Bell, 1973; Cacciari & Tabossi, 1988; Swinney & Cutler, 1979). The implication in all of these results is that regular forms take longer than irregular forms, implying that the regular forms must wait until the list of exceptions has been searched.

Note that these studies were highly controlled, closely matching the frequencies of the stems, the frequencies of the inflected forms, and the frequencies of the suffixes tested to ensure results were not confounded by these simpler frequency effects.
2.2.2 Maximize Productivity

A central aspect of the Tolerance Principle model is that it is intended as a kind of evaluation metric (Chomsky, 1965; Chomsky & Halle, 1968). Part of Chomsky’s (1965) model of language acquisition, the evaluation metric is a "method for selecting one of the (presumably infinitely many) hypotheses that are ... compatible with the given primary linguistic data." As such a metric, the Tolerance Principle allows the child to use the input – specifically, the number of words (N) to which a rule may apply and the number of those which are exceptions (e) – to determine the productivity of a morphological process. Importantly, Yang (2016) intends for the Tolerance Principle to be recursively applied. That is, it is intended that the child will continue to apply this evaluation metric to determine the productive status of rules as she accumulates evidence from her input over time.

To motivate this recursive evaluation, Yang (2016) assumes that learners are biased to "pursue rules that maximize productivity." That is, if no productive rule initially emerges for a full set of items (e.g., all verbs), the bias to maximize productivity motivates the learner to continue searching for a rule, perhaps applying the Tolerance Principle over some subset of the verbs (e.g., verbs with certain phonological features). This recursive application of the Tolerance Principle allows learners to detect "nested" rules, a common way in which natural languages organize rules. Yang (2016) illustrates the concept of nested rules using the English nominal suffixes -ness and -ity. While the -ness suffix appears to have no restrictions on the kinds of adjectives to which it may apply (the general rule), the -ity suffix applies productively only to adjectives ending in -ible, -ic, and -alis. The -ity suffix is thus considered to be a productive rule that is "nested" within the more general -ness productive rule for adjectives.
While it is not clear what kind of evidence would be required to confirm that learners have such a "maximize productivity" bias, one can use data from natural language to investigate whether the Tolerance Principle, when recursively applied, is indeed capable of finding these nested rules. For a number of cases in which one finds nested rules, including for the nominal suffixes *-ness* and *-ity* investigated above, the Tolerance Principle is indeed able to account for the productive status of rules (see Yang, 2016, for applications to English metrical stress, English nominalization suffixes, and German plural). Indeed, we will consider this application of the Tolerance Principle to explain the productivity of the low frequency *-s* suffix in the German plural system in Section 2.4.3.

2.2.3 **Optimize (Minimize) the time complexity of language use.**

Finally, the Tolerance Principle assumes that learners are biased to organize the regulars and irregulars such that the time complexity of language use is minimized (or, in other words, optimized). That is, learners will adopt the more computationally efficient strategy: store everything in a frequency ranked list, or store only the exceptions and form a productive rule. The idea that learners are biased toward more efficient strategies, or that processes are optimized to be maximally efficient is certainly not new. Indeed, such themes are pervasive in theories of linguistics, psychology, cognitive science, and neuroscience. As such, this assumption seems to be a reasonable one to make. Still, it remains to be seen whether learners indeed choose the optimal strategy between the two options proposed by the Tolerance Principle. And, even so, it is not clear whether the optimal strategy is constant across development, or whether the optimal strategy differs between learners in different maturational states (e.g., childhood vs. adulthood). We explore this latter point in greater detail in Chapter 6.
2.3 Formalization of the model

Formally, if \( R \) is a rule that may apply to \( N \) lexical items and there are \( e \) exceptions to this rule, the time required to access the rule can be expressed as \( T(N, e) \). If \( R \) is productive, as described in Section 2.1 (2), then the rule is not applied until the learner has first evaluated and rejected every exception (\( e \)) on the list. In other words, applying a productive rule consumes \( e \) units of time. The time required for exceptions, on the other hand, is determined by the lexical items frequency (i.e., its rank in the list of exceptions). To compute the time complexity \( T(N, e) \), Yang (2016, page 48) calculates "the weighted average of time units over the probabilities of these two sets of items." If \( R \) is unproductive, as in described in Section 2.1 (1), then all \( N \) items are treated as exceptions and are listed in order of frequency. The time complexity under these circumstances can be expressed \( T(N, N) \), as the number of exceptions \( e \) is equivalent to the number of items in the list \( N \). It is conjectured that the learner compares the time complexity required to form a productive rule, \( T(N, e) \), with the time complexity required when all \( N \) items are stored individually as lexical exceptions, \( T(N, N) \). By solving this equation for \( e \), the Tolerance Principle computes the precise number of exceptions that a productive rule can tolerate before its formation becomes computationally inefficient. The solution to this equation gives us the Tolerance Principle as follows:\(^2\):

(4) Tolerance Principle: Let \( R \) be a rule that is applicable to \( N \) items, of which \( e \) are exceptions. \( R \) is productive if and only if:

\[
e \leq \theta_N = \frac{N}{\ln(N)}
\]  \hspace{1cm} (2.1)

In other words, it is only more efficient to form a productive rule when the number of exceptions, \( e \), is less than the number of items, \( N \), divided by the natural log of the

\(^2\)The full derivation of the Tolerance Principle is available in the Appendix B
number of items. To illustrate, imagine a category of 9 items. Given a rule, R, that may apply to these 9 items, the Tolerance Principle predicts that 4.096 (or 9−9/\ln9) exceptions will be tolerated before forming a productive rule becomes less efficient than storing individual items. Thus, learners will form a productive rule if there are 4 or fewer exceptions to the rule R, but not if there are 5 or more. Importantly, this implies that the distinction between forming a productive rule and storing individual lexical items is a categorical one.

There is a theoretical tipping point at which forming a productive rule becomes less computationally efficient than the alternative strategy. The Tolerance Principle allows us to compute this tipping point.

2.4 Model performance on the common problems

The Tolerance Principle accurately predicts which rules are productive and which are restricted in dozens of languages. For example, using corpus analysis, Yang (2016) finds that there are 1022 unique past tense types in a large corpus of child-directed English.\(^3\) Thus, by the Tolerance Principle, the regular "add -d" rule for English past tense can tolerate \(\theta_{1022} = 1022/\ln(1022)\), or 147, exceptions or fewer. Of these 1022 past tense types, Yang (2016) finds that only 127 are irregular forms – well below the Tolerance Principle threshold of exceptions for productivity – indicating that the "add -d" rule is indeed productive. On the other hand, when Yang (2016) applies the same process to the irregular past tense rules, for example, sing/sang or fly/flew, a different result emerges. The same corpus yields 8 unique forms analogous to sing

\(^3\)Yang (2016) performed his calculations on multiple corpora, each yielding slightly different values for \(N\) and \(e\). However, the results – whether or not the Tolerance Principle accurately predicted productivity – were consistent across all corpora. Thus, for simplicity, we include numerical values only for his analysis of a single corpus, CHILDES, a five-million-word corpus of child-directed speech in English (MacWhinney, 2000).
(e.g., bring, fling, ring, spring, sting) and 22 unique forms analogous to fly (e.g., cry, lie, try, blow, grow, snow). Thus, the Tolerance Principle predicts that the sing/sang and fly/flew rules will be productive if there are less than or equal to 3.85 and 7.12 exceptions, respectively. When the number of exceptions to each rule are counted, these values are exceeded for both sing/sang, for which there are 5 exceptions, and fly/flew, for which there are 16.

How does the Tolerance Principle fare on the common problems reviewed in Section 1.4? In the following sections we review Yang’s (2016) application of the Tolerance Principle to these issues, turning first to the U-shaped development of English past tense.

2.4.1 The U-shaped function of past tense development

Recall that the Tolerance Principle is based on two critical values, N and e, where N is the number of items to which the rule may apply and e is the number of these N that are exceptions to the rule. Importantly, neither N nor e refer to all possible words to which a given rule may apply. Rather, N and e refer to the number of regular and irregular forms that a learner has acquired at a particular point in her acquisition. Further, recall that the Tolerance Principle is intended as an evaluation metric that is recursively applied to the values of N and e as the child accumulates more data. Thus, the Tolerance Principle is capable of precisely predicting the productive status of a rule for a given child, at any given point in his language development. However, to apply the Tolerance Principle in this manner, we must know the precise values of N and e for the individual child.

It is clearly impossible to know the precise number of regular and irregular forms an individual child has acquired at a particular point in time. However, Yang (2016) suggests that one way to estimate the likely values of N and e is to extract these
values from the top N most frequent items in child directed speech. For example, of the top 100 verbs in English (N=100), 54 are exceptions to the regular form (e=54). Of the top 500, 103 are exceptions. By considering a number of different values of N, which we can roughly assume correspond to different points in language acquisition, we can determine an estimate of the point at which a rule becomes productive. A summary of Yang’s (2016) results for this analysis on the English past tense "add -d" rule is available in Table 2.1. For children learning the English past tense, we can see that the number of exceptions remains too high to support the productivity of the "add -d" rule until the child’s vocabulary has become quite large (1022 past tense tokens).

Another way to estimate the number of regular and irregular forms a child has acquired is to approximate their vocabulary from the words they have been known to produce. Yang (2016) takes this approach to test the Tolerance Principle on the U-shaped development of past tense. The original empirical work demonstrating the U-shaped development of past tense (Marcus et al., 1992) used a corpus of child production data from a set of children who were recorded for long periods of time.

Table 2.1: Tolerance Principle threshold of productivity for the English past tense as vocabulary size increases over time. Adapted from Yang (2016).

<table>
<thead>
<tr>
<th>Top N</th>
<th>e</th>
<th>$\theta_N$</th>
<th>prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>54</td>
<td>22</td>
<td>unproductive</td>
</tr>
<tr>
<td>200</td>
<td>76</td>
<td>37</td>
<td>unproductive</td>
</tr>
<tr>
<td>300</td>
<td>92</td>
<td>52</td>
<td>unproductive</td>
</tr>
<tr>
<td>500</td>
<td>103</td>
<td>80</td>
<td>unproductive</td>
</tr>
<tr>
<td>800</td>
<td>121</td>
<td>119</td>
<td>unproductive</td>
</tr>
<tr>
<td><strong>1022</strong></td>
<td><strong>127</strong></td>
<td><strong>147</strong></td>
<td><strong>productive</strong></td>
</tr>
</tbody>
</table>
throughout their language development (Brown, 1973). Yang (2016) uses the data from these same children to determine whether their individual U-shaped functions can be predicted by the Tolerance Principle. To illustrate, I will focus on his analysis of one child, Adam (but see Yang (2016) for an analysis of other children). Using Adam’s data, Yang (2016) calculates the number of regular and irregular forms that Adam has produced up until the moment he produced his first overregularization ("what dat felt like?", age 2;11), as this is the first point at which one can be certain Adam has formed a productive rule 4. This approach yields 300 verbs, of which 57 are irregular. By the Tolerance Principle, Adam should tolerate only $\theta_{300} = 53$ exceptions. Yang (2016) calls this "agonizingly close" and attributes the difference to sampling effects. After this point, Adam may continue to make overgeneralizations of the regular form until he has learned and stored the correct irregular forms. How does Yang (2016) account for the early "correct period"? One possibility is that children have not yet fully stored these early correct forms. When the regular form becomes available, the memory for the irregular form is not yet strong enough, so the child resorts to overregularization. Only later and with more experience does the memory for the irregular form become strong enough to override the overregularization.

2.4.2 Error patterns and overregularization

As reviewed in Section 1.4.2, the evidence from the acquisition literature suggests that, while children make frequent overregularization errors, they almost never overgeneralize irregular forms. How does the Tolerance Principle account for this distinction?

4Yang (2016) acknowledges that this approach will likely underestimate the number of regular and irregular forms the child knows, as what is available in recordings of the children will not necessarily contain every regular and irregular the child has ever uttered (or has learned but has not yet uttered). Still, it provides a reasonable estimate with which to conduct the analysis.
Yang (2016) approaches this problem in the same way he approached the trajectory of the past tense over time. He used the top $N$ most frequent words to approximate young children’s vocabulary. When Yang (2016) applies this approach to the irregular patterns in English past tense, for example, *sing/sang* and *fly/flew*, the Tolerance Principle predicts that these patterns will be unproductive beginning at smaller values of $N$ (or estimated vocabulary size). The *sing/sang* pattern is predicted to be productive for the top 500 words\(^5\), but once the child’s vocabulary reaches 800 words, the pattern is no longer productive. This effect is even stronger for the *fly/flew* irregular form, which becomes and remains unproductive when the child’s vocabulary reaches only 200 words.

Further, Yang (2016) suggests that the Tolerance Principle can help explain variation in overgeneralization of the regular form in individual children. In a reanalysis of Brown (1973), Yang (2002) finds that the overregularization rates for Adam, Eve, Sarah, and Abe were 1.8%, 7.8%, 3.5%, and 24%, respectively. According to Yang (2016), these large variations, especially between Abe and the other children, could be accounted for by individual variations in vocabulary size and composition. For example, Yang (2016) points out that while Adam’s first overregularization came at age 2;11 ("what dat feel like?"), Abe was already overregularizing at his first recording session at age 2;3 ("he falled"). This observation suggests that Abe had learned the "add -d" productive rule much earlier than Adam. While the Tolerance Principle could not be applied to Abe, as there was no data available before his first overregularization, Yang (2016) finds that Abe’s vocabulary was indeed much

---

\(^5\)This suggests that there are very early points at which children may have a productive *sing/sang* rule. However, Yang (2016) points out that these early points are likely before the child is capable of producing any past tense inflections at all. And even if they can, this might help explain the somewhat more frequent overirregularization of "brang" in the acquisition data.
larger than Adam’s, suggesting further that a child’s individual vocabulary size is an important variable in determining whether or not a process is productive.

Finally, recall that the literature reviewed in Section 1.4.2 suggested that the majority of errors that young children make during language acquisition are errors of omission. Regarding these errors, the Tolerance Principle makes a similar prediction to Pinker (1999). Before children have acquired a productive rule, they will be unable to inflect novel forms they have not yet memorized and will be forced to produce an omission error (the uninfected stem). However, the Tolerance Principle provides a model with which one can determine precisely when the child should or should not have formed a productive rule, and therefore, should or should not produce errors of omission.

While the Tolerance Principle has not been applied to specifically this aspect of children’s errors beyond what is discussed in the previous paragraph, Yang (2016) does apply the Tolerance Principle to a similar phenomenon in natural language: paradigmatic gaps. Recall from Section 1.4.2 that paradigmatic gaps are arbitrary gaps in the lexicon for which there is no productive or default form (Halle, 1973), and that these gaps occur fairly commonly in natural language. Yang (2016) uses the Tolerance Principle to demonstrate that, after recursive applications to locate possible nested rules, no productive forms emerge in these cases. That is, each case has too many exceptions to the rule to survive the Tolerance Principle threshold for productivity. For example, in English, "stride" is known to be an irregular verb, as its past tense form is "strode." However, no past participle form exists. When Yang (2016) analyzed the possible past participle patterns that could apply to *stride* — *strode* and *stridden* — it turned out the neither survived the Tolerance Principle threshold for productivity. For example, in English, "stride" is known to be an irregular verb, as its past tense form is "strode." However, no past participle form exists. When Yang (2016) analyzed the possible past participle patterns that could apply to *stride* — *strode* and *stridden* — it turned out the neither survived the Tolerance Principle threshold for productivity.
threshold for productivity. For *strode*, Yang (2016) found that 52 of the 95 irregular verbs have the same past and past participle form, while 43 are exceptions. By the Tolerance Principle, this rule is productive only if there are 21 or fewer exceptions. For *stridden*, 14 verbs share the structural description analogous to *ride-rode-ridden*, resulting in a Tolerance Principle threshold of 5 or fewer exceptions. Of these 14 verbs, 6 are exceptions. Thus, as predicted, neither *strode* nor *stridden* meet the threshold for productivity.

2.4.3 German plural

As in the previous example of paradigmatic gaps, the Tolerance Principle accounts for the productivity of the low-frequency German plural -s by applying recursively to locate increasingly more specific, nested rules. When the entire scope of German nouns is considered, none of the five possible plural suffixes (-(e)n, -e, -er, -s, and null) survive the Tolerance Principle threshold for productivity. However, according to the Tolerance Principle’s Maximize Productivity assumption, the failure of a productive rule to emerge will motivate the learner to search for productive rules that may apply to some subset of the full set of nouns, thus revealing nested rules. Indeed, by applying the Tolerance Principle to increasingly more specific structural generalizations (e.g., all feminine nouns, all non-feminine nouns ending in a reduce syllable, etc.), Yang (2016) reveals a nested structure of productive rules from which the low frequency -s suffix emerges as default. These nested rules are as illustrated in 5.

(5) a. If feminine, then add -(e)n
   b. If not feminine, then:
      i. If ends in reduced syllable, then add null (-∅)
      ii. If ends in schwa, then add -n
iii. If monosyllabic, neuter, and contains back vowel, then add \(-er\)

iv. Otherwise, add \(-e\)

c. Otherwise, add \(-s\)

To discover this nested rule structure, Yang (2016) began by broadly dividing the nouns into those that are feminine and those that are not. Of the 166 feminine nouns in his sample of the 500 most frequent nouns in a corpus of child-directed German (MacWhinney, 2000), 146 take the \(-en\) suffix and 20 are exceptions to this rule. According to the Tolerance Principle threshold for productivity, the "add \(-en\)" rule can tolerate as many as 32 exceptions. Thus the Tolerance Principle predicts that "add \(-en\)" is productive for feminine nouns. Turning next to the non-feminine nouns, further nested rules emerged based on more specific phonological regularities. For example, in the same corpus, of the 83 non-feminine nouns that end in a reduced syllable, 77 take the null \(-\) suffix and 6 are exceptions. Again, according to the Tolerance Principle threshold for productivity the "add \(-\)" rule can tolerate 18 exceptions. Thus "add \(-\)" is productive for non-feminine nouns that end in a reduced syllable. The same result applies to the nested rules described in 5.b.ii. and 5.b.iii.

When the Tolerance Principle is applied to only the subset of nouns that share the same features, there are not enough exceptions to overturn these productive rules. For rule 5.b.iv., if the noun is not feminine, and the more specific productive rules above it do not apply, then the noun is inflected by \(-e\). Finally, if none of the productive rules in a and b apply to a particular noun, the "add \(-s\)" productive rule is applied.

According to Yang (2016) and the Tolerance Principle, the German plural \(-s\) suffix survives as the default form for two specific reasons. First, because these rules are nested, nouns that can be described by the more specific rules do not count against the productivity of "add \(-s\)" itself. Instead, they apply only to their specific subset
of nouns and are exceptions only to their specific productive rule (e.g., non-feminine nouns that end in schwa either take the productive rule -n, or are exceptions to this rule). Second, while the other productive suffixes apply to nouns with specific features (e.g., gender), the -s suffix is not similarly restricted. Thus, because it applies to the most general set of nouns after the more specific productive rules have been traversed, it emerges as the default productive form.

2.5 The need for converging evidence from child psycholinguistics

As we have seen in the previous section, the Tolerance Principle has been tested on a wide range of empirical phenomenon, from overgeneralization patterns in language acquisition to the problem of German plural. While this evidence alone is quite convincing, subjecting the Tolerance Principle to further empirical scrutiny via artificial language learning experiments could provide further tests of the efficacy of the Tolerance Principle as a model of productive rule formation. In particular, as we have previously pointed out, corpus analyses – even those performed on the productions of individual children – can only provide estimates of the values of N and e. They are unlikely to provide a true sample of all the regular and irregular forms a given child has learned. On the other hand, artificial language learning experiments allow us precise control over the language that children are exposed to. We can control precisely how many regulars and how many exceptions the child acquires and determine whether children obey the Tolerance Principle in their formation of productive rules.

Further, artificial language experiments allow us complete control over every feature of the language, including the type and token frequencies of the forms children are exposed to. Thus, we can specifically manipulate these variables in order to ask how the Tolerance Principle withstands variation across type and token frequency
that are present in the input. We can also use these artificial languages to compare the Tolerance Principle directly to other metrics of productivity to determine which models most closely approximate the behavior of children.

Finally, by conducting artificial language learning experiments in both children and adults, we can ask whether the Tolerance Principle is a model that is exclusive to children, or whether this model applies to the acquisition of productive rules in language regardless of the maturational state of the learner.

It is important to note here that it is not the goal of this dissertation to argue that the computations underlying the Tolerance Principle are reflecting exactly the computations that a child performs to determine whether a rule is productive or not. Rather, that the Tolerance Principle model is capturing something crucial about the categorical nature of children’s behavior under these circumstances, and that whether there is a productive rule or not appears to be correlated with the number of lexical items that follow the rule, as predicted by the Tolerance Principle.

2.6 Overview of experiments

The primary goal of this dissertation is to test the Tolerance Principle as a model of productivity through a series of artificial language learning experiments in children and adults. In Experiment 1 (Chapter 3) we provide a simple first test of the Tolerance Principle in an artificial language in which the token frequency of the regular form is high. In Experiment 2 (Chapter 4), we adjust the token frequency of the regular form such that the distribution of exceptions is more like that of natural language. In Experiment 3 (Chapter 5), we compare the Tolerance Principle to another evaluation metric, a simple majority of types metric, to determine which of the two metrics children prefer to follow. Further, throughout these three experiments, we test both
children and adult participants in order to address whether the Tolerance Principle is a model of acquisition that is exclusive to children. In Experiments 4 and 5 (Chapter 7), we explore this concept in more detail, adding additional experiments with adults to attempt to understand the differences we observe between children and adults across our three experiments.
Chapter 3

Testing the Tolerance Principle in children learning an artificial language

3.1 Introduction to Experiment 1

The primary goal of Experiment 1 was to provide a first test of the Tolerance Principle in children acquiring an artificial language. To this end, we created a simple artificial language containing 9 nonsense nouns and used the Tolerance Principle to calculate the number of exceptions a productive rule should tolerate under these circumstances ($\theta_N \leq 4.096$). Then, we exposed children to one of two language conditions: one in which the Tolerance Principle predicts a productive rule should be formed (5 regulars, 4 exceptions), and one in which a productive rule should not be formed (3 regulars, 6 exceptions). Following this exposure, we assessed whether children had formed a productive rule by determining whether or not they applied the rule to novel nouns in a production test. A detailed description of these experimental methods is provided in Section 3.2.2.

We had two additional goals in mind for Experiment 1. First, recall that the Tolerance Principle is based on the number of types that take the regular form. That is, how many exceptions there are ($e$) to a regular form in a given category of items ($N$). The token frequency, or the raw number of times an item appears in a corpus,

---

1This experiment appeared in Schuler et al. (2016)
is not necessarily relevant to whether children following the Tolerance Principle will or will not form a productive rule.

The idea of productivity based on type and not token frequency is not necessarily disputed. In fact, most psychological and linguistic theories of regular and irregular morphology accept the elevated importance of type over token frequency in the formation of productive rules (e.g. Baayen & Renouf, 1996; Bybee, 1995; Pinker, 1991; Plunkett & Marchman, 1991; Rumelhart & McClelland, 1986). Nevertheless, differentiating between type and token frequency is especially relevant in our artificial language experiment given previous findings in the artificial language learning literature. Specifically, several experiments have demonstrated that when learners are faced with inconsistent input — that is, types that are inconsistently or probabilistically marked across tokens — adult learners will closely track token frequencies during exposure and will match these frequencies in their productions (Austin, 2010; Hudson Kam & Newport, 2005, 2009). While children in these experiments do not similarly match the token frequencies in their productions, they behave in systematic ways based on the token frequency. That is, they regularize the majority token in their input, applying this form significantly more frequently than it occurred in the input (Austin, 2010; Hudson Kam & Newport, 2005; Singleton & Newport, 2004). Often this regularization behavior is so robust that children in these experiments will apply the majority token categorically in their own productions, using this form on 100% of production test trials.

While there are a number of differences between these experiments and the present work — most notably that every type is probabilistically marked in the previous inves-

---

2Note, however, that the token frequency of a word is not completely irrelevant to the Tolerance Principle. The time complexities from which the Tolerance Principle is derived make use of the Zipfian distribution of word frequencies. However, the final solution requires only that child know the number of types (N) and how many of those are exceptions.
tigations, while here each type is marked with a consistent form – it is important to determine whether children are tracking type frequency in the present experiment as the Tolerance Principle suggests, or whether they are tracking token frequencies and simply regularizing the token that appears most often in the language input, as children in similar artificial language experiments have been shown to do.

In Experiment 1, we investigate the type-token distinction by pitting type and token frequencies against one another. To do so, we keep the token frequency of the regular form high across all conditions, such that children tracking this statistic should form a productive rule regardless of the language condition they participate in. At the same time, the type frequencies in our experiment are such that children following the Tolerance Principle should form a productive rule in one condition (5 regulars and 4 exceptions) and should not from a productive rule in the other (3 regulars and 6 exceptions).

Our second additional goal was to begin to address whether the Tolerance Principle is an model of productive rule formation that is unique to children, or whether this model predicts adult behavior as well. According to Yang (2016), the Tolerance Principle is a model in which children recursively apply this evaluation metric in order to determine the status of a rule (productive or not) as they accumulate evidence from their input. Though Yang does not explicitly exclude the possibility that adults employ the Tolerance Principle, he suggests that children may be in better position to acquire rules from this approach based on a "smaller is better" hypothesis; that is, the proportion of exceptions a productive rule can tolerate dramatically decreases as the number of items over which the rule may apply \((N)\) increases. Given this consequence of the Tolerance Principle, Yang reasons that learners with smaller vocabularies may have an easier time learning a rule, and thus children may have a serious advantage over adults. This "smaller is better" hypothesis is reminiscent of the "Less is More"
theory originally proposed by Newport (1990). The "Less is More" hypothesis suggested that children are better language learners precisely because of their "limited cognitive capacity." Importantly, a number of studies both preceding (e.g. Curtiss, 1977; Johnson & Newport, 1989) and following (e.g. Mayberry & Eichen, 1991; Perani et al., 1996; Hudson Kam & Newport, 2005; Culbertson & Newport, 2015, e.g.) Newport’s proposal lend support to this theory, finding striking differences between the ways in which children and adults acquire languages. Perhaps the Tolerance Principle captured what children, but not adults, do during the language acquisition process. We explore this possibility by including adult participants in this and all following experiments.

3.2 Experiment 1 materials and methods

3.2.1 Participants

Fifteen children (mean = 7.55 years, standard deviation = 0.86 years, range = 6.05 – 8.89 years) and twenty adult controls participated in this experiment. An additional three children began the experiment but did not complete it, and an additional four children participated but were excluded from analysis for failure to understand the task (quantified as a failure to produce the correct noun on at least 50% of the test trials). Children were recruited from schools, camps, and day-cares in the Washington D.C. metro area and participated in the experiment either at these off-site locations or in our lab. Adult control participants were recruited and run on-line via Amazon Mechanical Turk.

All participants were native English speakers with normal hearing and normal to corrected-to-normal vision. Child participants received stickers and a set of small toys for their participation. Parents of children who participated in our lab received
$10 in additional compensation for travel reimbursement. Adult participants received compensation at a rate of $10 per hour.

3.2.2 Stimuli

To assess whether children learning an artificial language follow the Tolerance Principle, we designed two artificial language conditions: one in which the Tolerance Principle predicts that learners should form a productive rule and one in which learners should not form a productive rule. To do so, we first created a rule $R$ for a category of 9 nonsense nouns. The rule was: To make a noun plural, add ka. Next, we used the Tolerance Principle to calculate the number of exceptions a productive rule can tolerate in a category of 9 nouns. Using the predicted value of 4.096 exceptions, we created two conditions: one where a productive rule should be formed (5R4E: 5 regulars, 4 exceptions), and one in which the Tolerance Principle predicts a productive rule should not be formed (3R6E: 3 regulars, 6 exceptions). That is, children should tolerate 4.096 exceptions or fewer in a category of 9 nouns. In the 5R4E condition, there are only 4 exceptions and children should, therefore, form a productive rule. In the 3R6E condition, there are 6 exceptions and children should not form a productive rule.

To create our exposure corpus, we assigned each noun a plural marker that either followed the rule (add $ka$) or was an exception (add $po$, $tay$, $lee bae$, $muy$, or $woo$), depending on the condition. Then we used these nouns and markers to create an exposure corpus of 72 sentences (24 singular and 48 plural). All sentences began with the same nonsense verb $gentif$, meaning "there is/are". Singular sentences were unmarked ($gentif + NOUN$) and paired with one image of the corresponding object. Plural sentences were formed $gentif + NOUN + MARKER$ and paired with 2, 4, or
Table 3.1: Frequency of nouns and their plural markers in Experiment 1, Language A. The regular form appears in bold.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Frequency</th>
<th>N Plural</th>
<th>Noun</th>
<th>5R4E</th>
<th>3R6E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>16</td>
<td>mawg</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>8</td>
<td>tomber</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>5</td>
<td>glim</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4</td>
<td>zup</td>
<td>ka</td>
<td>po</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>4</td>
<td>spad</td>
<td>ka</td>
<td>lee</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>3</td>
<td>daygin</td>
<td>po</td>
<td>bae</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3</td>
<td>flairb</td>
<td>lee</td>
<td>tay</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>3</td>
<td>klidam</td>
<td>bae</td>
<td>muy</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>lepal</td>
<td>tay</td>
<td>woo</td>
</tr>
</tbody>
</table>

6 images of the corresponding object (see Figure 3.1). There were thus 18 possible sentences in the language: 1 singular and 1 plural sentence for each noun.

We generated the exposure corpus by allowing noun frequency to vary approximately along a Zipfian distribution (Zipf, 1949), with nouns taking the regular form, *ka*, as the most frequent in both conditions. Thus the second most frequent noun was presented about half as often as the most frequent noun, the third most frequent noun was presented about half as often as the second, and so on. This distribution is important to our experimental design for at least two reasons. First, the distribution of word frequency in natural language is approximately Zipfian, and the computations underlying the Tolerance Principle assumes that word frequencies follow this pattern. Thus, to rigorously test the Tolerance Principle, we must ensure that word frequencies in our artificial language follows this pattern as well. Second, has previously been mentioned, the Tolerance Principle is computed based on the number of types that take the regular form, regardless of the raw token frequency of this form.
in the input. This computation suggests that children are keeping track of type and not token frequency as they are acquiring language. By applying the regular form to the most frequent nouns in our Zipfian distribution, we create a situation in which type and token frequencies are in conflict. If children follow the Tolerance Principle, tracking the number of types that take the regular form, they should form a productive rule in the 5R4E condition, but not in the 3R6E condition. However, if they are instead tracking the token frequency of the ka marker in their input, they should regularize the regular form in both 5R4E and 3R6E, forming a productive rule in both experimental conditions.

Finally, to ensure that there was nothing idiosyncratic about the particular noun frequency ranking, we created two counterbalanced language: A and B. Language A and B differed only in that, in language A, mawg was the most frequent nouns and lepal was the least frequent noun, while in language B, this noun ranking was reversed (lepal was most frequent and mawg least).

3.2.3 Procedure

The experiment consisted of three parts: exposure to the language, a production test, and a rating test. During exposure, participants were presented with the 72 exposure sentence-image pairs in random order. On each trial, participants saw a picture of 1, 2, 4, or 6 instances of a noun and were presented with the corresponding singular (for 1) or plural (for 2, 4, or 6) sentence (see Figure 3.1a). Children heard these sentences under high quality headphones while Mechanical Turk participants saw these sentences written under the corresponding image in their browser window. Participants were asked to repeat the sentence aloud (or type it into a response box.
on Mechanical Turk) before moving on to the next trial. Every 18 trials they were given a short break. Children were offered a sticker during breaks to encourage them to continue in the experiment.

![Image](image.png)

(a) Four example exposure trials  
(b) Example production test trial

Figure 3.1: **Examples trials for exposure phase and production test in our experimental paradigm.** The same paradigm was used in all experiments (Experiments 1–5).

After exposure, we used a production test modeled closely after the "wug" test to assess whether children had formed a productive rule (Berko, 1958). During this test, participants were given singular sentence-image pairs containing novel nonsense nouns they had not heard during exposure and were asked to provide the plural form (see Figure 3.1). Each participant completed 12 production test trials, 2 for each of 6 novel nouns (bleggin, daffin, norg, sep, fluggit, and geed). To prevent participants from using a plural form based only on the precise number of instances shown in that trial, the test items contained 3 or 5 instances of the novel noun (whereas there were 2, 4, or 6 in the exposure set).

Following the production test, we gave participants a rating test to ensure they had learned the nouns and markers they were exposed to. Because the number of

---

3We have conducted many experiments using this paradigm, both in our lab and on Mechanical Turk. We find no significant difference in adult performance between the in-lab procedure and the slightly modified Mechanical Turk procedure.
nouns taking the regular form in our exposure is not equal across conditions, we sampled only a subset of nouns for the rating test such that the test was balanced across conditions. We sampled a total of four nouns from the exposure set, two that followed the regular form and two that were exceptions. Due to the distribution of the regular nouns in the frequency rank, this also corresponded with two high-frequency nouns (rank 1 and 2 in the Zipfian distribution) and two low-frequency noun (rank 7 and 9 in the Zipfian distribution). Each of these four nouns appeared four times in the rating test; twice with the correct marker and twice with an incorrect marker, resulting in a total of 16 rating test trials.

3.3 Experiment 1 results

For each production test trial, participants were asked to produce the plural form of a novel noun after hearing this new noun in its singular form. These novel productions allowed us to assess whether participants formed a productive rule. Recall that the Tolerance Principle predicts that there will be a categorical distinction between productive and unproductive (lexically specific) rules. In our artificial language, a productive rule should be formed if more than 4.096 nouns obey the rule (as in our 5 regular/4 exception condition), but not if fewer than 4.096 nouns do (as in our 3 regular/6 exception condition). When a productive rule is formed, it should be applied to 100% of novel nouns, as is the case for English past tense 'add ed.'

To determine whether participants formed a productive rule, we performed a one-tailed t-test against the hypothesized value of 100%. Participants who have formed a productive rule should, according to the Tolerance Principle, mark these novel plural sentences with ka 100% of the time. On the other hand, participants who have not
formed a productive rule should use the *ka* marker significantly less than 100% (and perhaps no more frequently than any other plural marker is used).

Focusing first on the child production data, Figure 3.2 shows the percentage that each inflection type was produced during the production test for participants in the 5 regular/4 exception condition (5R4E) and the 3 regular/6 exception condition (3R6E). These data show that children in the 5R4E condition mark novel nouns with the *ka* inflection on 91.7% of plural trials. This value is not statistically different from 100% (*t*=1.00, *p*=0.18). In contrast, in the 3R6E condition, children mark novel nouns with the *ka* inflection on only 16.9% of plural trials. This value is substantially and significantly different from 100% (*t*=6.81, *p*<0.0001). Children thus appear to have
formed a productive rule when the Tolerance Principle predicts that they will (in the 5 regular/4 exception condition), but not when it predicts that they will not (in the 3 regular/6 exception condition).

This strong result is further underlined by looking at the data from individual children. Figure 3.3 shows a dot plot of individual children’s use of the ka inflection in the 5 regular/4 exception condition compared with the 3 regular/6 exception condition. Importantly, six out of 7 children produced the ka inflection on 100% of production trials in the 5 regular/4 exception condition, while only one out of 8 children did so in the 3 regular/6 exception condition.

Turning next to the adult production data, we find a somewhat different pattern of results. Adults in the 5R4E condition mark novel nouns with the ka inflection on
65.0% of plural trials; unlike children, this value is significantly different from our 100% productivity criterion ($t=3.23, p<0.01$). Like children, adults in the 3R6E mark novel nouns with ka significantly less than 100% of plural trials ($t=4.59, p<0.001$) - only 51.7%. That is, for adults, this contrast is much weaker, and when we compare the use of the ka inflection between adults to the children, as in Figure 2, we see striking differences between the two. The Tolerance Principle effect is much more pronounced in children, who exhibit a much more categorical response in their use of the ka inflection. Indeed, children, but not adults, show a significant difference between the use of ka in the 5R4E condition and that in the 3R6E condition (children: $t=4.91, p<0.001$, adults: $t=0.89, p=0.39$).

One possible explanation is that adults are not obeying the Tolerance Principle and are instead producing ka with the same frequency with which this marker occurred in their input. This behavior is known as probability matching (Hudson Kam & Newport, 2009, 2005). Recall that the nouns in our artificial language follow a Zipfian distribution, with rule-following nouns being the most frequent in the distribution. Thus the token frequency of the ka inflection is fairly high in both conditions: 75% of the plural exposure sentences in the 5R4E condition, and 58.3% of plural exposure sentences in the 3R6E condition. To determine whether this could explain the difference between the two groups, we analyzed both child and adult use of ka against the token frequency of ka in the exposure for the two conditions. We found that only adults match the token frequency in both the 5R4E ($t=0.92, p=0.19$) and 3R6E conditions ($t=0.63, p=0.27$). In contrast, the child data is not consistent with a probability matching interpretation. Children in the 5R4E condition produce the ka inflection significantly more than the input frequency ($t=2.00, p<0.05$) and in the 3R6E condition produce the ka inflection significantly less than the input frequency ($t=3.40, p<0.01$).
Figure 3.4: **Usage of the regular form by children and adults compared with predictions made by the Tolerance Principle and the token frequency in Experiment 1.** Blue bars are the actual from child and adult participants. Red and gray bars are predicted data from the Tolerance Principle and token frequency, respectively. Error bars are standard error.
3.4 EXPERIMENT 1 DISCUSSION

In Experiment 1, we set out to determine whether the Tolerance Principle could accurately predict when children would and would not form a productive rule in an artificial language learning experiment. For our 9-noun artificial category, the Tolerance Principle predicts that a productive rule will be formed when there are fewer than 4.096 exceptions to the rule. We found that, just as predicted, children formed a productive rule when there were 4 lexical items that were exceptions to the rule, but not when there were 6. Combined with the existing analysis of natural language corpora (Yang, 2016), these results suggest that the Tolerance Principle has captured something significant about the conditions for productivity during rule acquisition.

Importantly, the criterion we used to assess whether children formed a productive rule was categorical. Our analysis asked whether learners extended the rule to 100% of the test trials - the most rigorous possible test of productivity. We found that, while both children and adults were more likely to extend a productive rule when there were 4 exceptions than when there were 6, only children displayed a categorical distinction between forming a productive rule and not forming one. As predicted by the Tolerance Principle, almost every child exposed to 5 regulars and 4 exceptions extended the rule to 100% of test trials, while almost no children exposed to 3 regulars and 6 exceptions did (see Fig 3.3). While children in the 3R6E condition did occasionally use ka on novel items, their results are only what one would expect under conditions of uncertainty. In the absence of a productive rule, children appear to be unsure how to mark novel forms, most often using no plural marker (null) and otherwise selecting at random from among the various markers they heard during exposure (the ka inflection as well as the exceptions).
The striking difference we observed between children and adults, shown in Figure 3.4, led us to ask whether an alternative model would better predict adult behavior in this task. As was mentioned in 3.1, the Tolerance Principle is based on the number of lexical items (types) that observe or violate a pattern. By contrast, adults in Experiment 1 appear to follow the token frequencies of the regular form more closely. Recall that, in order to disambiguate between type and token frequency in this experiment, we created our artificial language such that the token frequency of the ka marker was high in both the 5R4E and the 3R6E conditions (75% and 58.3%, respectively), while the type frequency was sufficient for productive rule formation in only one condition (5R4E). We found that, in line with previous artificial language studies on inconsistent input (e.g. Austin, 2010; Hudson Kam & Newport, 2005, 2009), adults produced the ka marker with the same frequency that this marker was presented in their input, suggest that adults adopt a similar strategy in performing these two different tasks.

Children, on the other hand, appear to handle these tasks quite differently. As previously discussed, in the literature on inconsistent input, children produce the form that appears most often in their exposure nearly 100% of the time (Austin, 2010; Hudson Kam & Newport, 2005). While the results of our 5R4E condition are similar to these findings, the results from our 3R6E condition are quite different. Although the ka marker is the most frequent marker in the 3R6E exposure (58% of tokens), children produced this form with much lower probability (16.9% of trials). This suggests that children are not simply forming productive rules based on what appears most frequently in their input, as they do when they are faced with patterns that are inconsistently marked or probabilistic. In the present paradigm, in contrast with the previous studies of inconsistent input, each lexical item is marked in a consistent way across trials. When children are exposed to lexically consistent patterns such as this,
they form productive rules based on the number of lexical items that observe these patterns, as predicted by the Tolerance Principle.

Finally, in Experiment 1, we included an adult control population in order to begin to determine whether the Tolerance Principle is model of productive rule formation that is special for children. Indeed, as predicted, our results are consistent with children following the Tolerance Principle and adults matching the token frequency. As discussed in Section 3.1, it was proposed that children may have an advantage in acquiring rules via the Tolerance Principle because they begin with very small vocabularies; a "smaller is better" hypothesis. However, our experimental results are unable to confirm whether this is, in fact, the case. That is, both children and adults in our artificial language experiment were exposed to exactly nine nouns and thus presumably had exactly the same vocabulary size in this rule learning scenario. While our results provide further evidence for differences between children and adults – children are clearly following the Tolerance Principle and adults are clearly not – it is not yet clear why children and adults behave so differently when faced with the same rule learning scenario.

One possibility is that probability matching may be the more efficient computational strategy for adult learners. This would imply that the Tolerance Principle is exclusive to children, capturing a basic principle of generalization in rule formation for very young learners. A related possibility is that only children learn or produce forms categorically. On this interpretation, the difference between 5R/4E and 3R/6E influences both child and adults learners, but only children show this difference in such an extreme contrast in their output. A final possibility is that perhaps adults only engage probability-matching behavior when they are learning from a very small number of items; probability matching may be easy and efficient for adults when there are only a few items to keep track of. However, when there are many items to
track, it may no longer be efficient to track and closely match input probabilities. This latter interpretation is supported by the results in studies on a different topic, the effect of inconsistent input. In these studies, children and adults look very different in experiments when the learning task involves only a small number of items, with adults probability matching and children maximizing the majority form in their productions. However, when the number of lexical items and their variations become very large, adults begin to behave more like children (Hudson Kam & Newport, 2005, 2009). In Experiment 4, we will explore this possibility further by giving adult participants our Tolerance Principle tasks with a very large set of nouns. We hypothesize that, as the number of nouns increases, adults may begin to exhibit the type of categorical behavior predicted by the Tolerance Principle.
4.1 Introduction to Experiment 2

In Experiment 1, we demonstrated that the Tolerance Principle accurately predicts when children will and will not form a productive rule in a highly controlled artificial language experiment. Further, by applying the regular form to the most frequent nouns in our Zipfian distribution, effectively pitting type and token frequencies against each other, we confirmed that children do, in fact, attend to the number of types that take the regular form and not token frequency when acquiring productive rules. However, one might question whether the artificial language in Experiment 1, with the regular form applied to the most frequent nouns, is especially ecologically (or typologically) valid. That is, is it common in natural language typology to have the regular form applied to the high-frequency items while the exceptions are applied only to low-frequency items?

One frequently encountered argument is that exceptions must be highly frequent in order to be learned and maintained in the language. That is, if exceptions were infrequent, these exceptional forms would be forgotten, inevitably regularized out of the language over time (e.g., Bybee, 2006; Corbett et al., 2001). In support of this hypothesis, a number of investigators have found that low-frequency irregulars are more likely to be accidentally over-regularized by both children (Marcus et al., 1992).
and adults (Bybee & Slobin, 1982; Stemberger, 1983), that less frequent an irregular the faster its rate of extinction in language evolution terms (Lieberman et al., 2007), and that, in general, exceptions have are biased to be high in frequency (Bybee, 1985).

Importantly, however, permitting some exceptions in the high end of the frequency distribution does not necessarily mean that the majority of the highest frequency items are exceptions, nor that all exceptions are highly frequent. The former is often assumed in the psychology literature, likely because of the unusual distribution of irregulars in English past tense (the most highly investigated case in language acquisition). While it is indeed true that the majority of the highest frequency verbs in child-directed English are irregular (54 of the top 100), if we expand our vocabulary a bit the regular form -ed turns out to apply to fully 86% of the top 1000 most frequent words1 (Pinker, 1999). Further, as Yang (2016, page 65) points out, "the concentration of English irregular verbs at the very top appears to be an outlier in a wide range of empirical cases." Among those wide-ranging cases is English plural, in which only 6 of the top 100 nouns are irregular (e.g., people, women, children).

Thus, to construct an artificial language that is ecologically valid, we can assume that our exceptions should, at most, be well distributed with the regular form among the high-frequency items. We can not, as one might have originally suspected, assume that exceptions should apply to all of the highest frequency items in the Zipfian distribution (effectively the opposite of Experiment 1). As we have shown above, a distribution of exceptions in which exceptions account for all of the highest frequency items is not empirically supported by the existing literature. Given these findings, we repeated Experiment 1 here with a modified (more ecologically valid) artificial language. In this new language, we altered which plural markers applied to which

1If we expanded our vocabulary even more, we would find that only 3% of all English verbs are irregular, making exceptions the overwhelming minority.
nouns in the frequency rank, such that the regular form and the exceptions were evenly distributed among the high-frequency items – a distribution more consistent with that of natural languages.

Given that Yang (2016) has demonstrated the effectiveness of the Tolerance Principle across dozens of natural language situations, including those in which exceptions are thought to be highly frequent (e.g., the infamous English past tense) and those for which the rule structure is highly complex (e.g., German plural)\(^2\), we hypothesize that children in Experiment 2 will continue to follow the Tolerance Principle, even though the regular form is no longer applied to the most frequent nouns. We test this hypothesis in the following experiment. As in experiment 1, we include a population of adult controls in order to observe whether children and adults behave differently in their acquisition of productive rules.

4.2 EXPERIMENT 2 MATERIALS AND METHODS

4.2.1 PARTICIPANTS

Twenty-one children (mean age = 6.7 years, standard deviation = 0.80 years, range = 5.19 – 7.82 years) and twenty-two adult controls (mean age = 20.72, standard deviation = 2.17 years, range = 18.05 – 25.43 years) participated in this experiment. An additional four children and six adults participated but were excluded from analysis for failure to meet experiment inclusion criteria (one child and one adult were non-native English speakers), failure to understand the task (quantified as failure to produce the correct noun on at least 50% of test trials; three children failed this criterion), or equipment malfunction (five adults experienced malfunction). Children were recruited from schools, camps, and day-cares in the Washington D.C. metro area and

\(^2\)For a review of the application of the Tolerance Principle to English past tense and German plural acquisition, see Section 2.4
participated in the experiment either at these off-site locations or in our lab. Adult participants were recruited from the Georgetown University community and were run in our lab.

All participants were native English speakers with normal hearing and normal to corrected-to-normal vision. Child participants received stickers and a set of small toys for their participation. Parents of children who participated in our lab received a $10 travel reimbursement as additional compensation. Adult participants received compensation at a rate of $10/hour.

4.2.2 Stimuli

The artificial language for Experiment 2 was designed in the same way as Language A in Experiment 1 (described in Section 3.2.2). However, in order to create a language in which regulars and exceptions were evenly distributed along our Zipfian distribution, we modified how these plural markers were applied to nouns in the frequency rank. In Language A, we evenly distributed the regulars and exceptions along the frequency rank, with the regular form applied to the most frequent noun, an exception applied to the second most frequent noun, and so on. To ensure that our results would not be driven by the fact that the regular from was applied to the most frequent noun, in Language B we again evenly distributed the regulars and exceptions along the frequency distribution, beginning instead with an exception applied to the most frequent noun. (see Table 4.1).

The production test remained unchanged from Experiment 1, testing children two times each on a total of six novel nouns (bleggin, daffin, norg, sep, fluggit, and geed).

The rating was modified in an effort to sample each noun once (rather than sampling only four of the nouns as in Experiment 1). Further, in order to simplify the test for the children, instead of performing a rating test on a 5-point scale, children
Table 4.1: Frequency of nouns and their plural markers in Experiment 2, Languages A and B.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Frequency</th>
<th>N Plural</th>
<th>Noun</th>
<th>5R4E</th>
<th>3R6E</th>
<th>5R4E</th>
<th>3R6E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>16</td>
<td>mawg</td>
<td>ka</td>
<td>ka</td>
<td>po</td>
<td>po</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>8</td>
<td>tomber</td>
<td>po</td>
<td>po</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>5</td>
<td>glim</td>
<td>ka</td>
<td>ka</td>
<td>lee</td>
<td>lee</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4</td>
<td>zup</td>
<td>lee</td>
<td>po</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>4</td>
<td>spad</td>
<td>ka</td>
<td>ka</td>
<td>bae</td>
<td>bae</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>3</td>
<td>daygin</td>
<td>bae</td>
<td>bae</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3</td>
<td>flairb</td>
<td>ka</td>
<td>tay</td>
<td>tay</td>
<td>tay</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>3</td>
<td>klidam</td>
<td>tay</td>
<td>muy</td>
<td>ka</td>
<td>muy</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>lepal</td>
<td>ka</td>
<td>woo</td>
<td>ka</td>
<td>woo</td>
</tr>
</tbody>
</table>

were asked for a binary judgment about "whether this sentence is correct." Importantly, every test trial sentences included the noun marked with the regular form (ka), requiring children to simply have learned whether a given noun takes the regular form or is an exception, and allowing us to keep the test consistent across conditions.

4.2.3 Procedure

All aspects of the procedure, including the exposure, production test, and rating test were performed in exactly the same way as described in Experiment 1 (see 3.2). However, note that adults in Experiment 2 were run in the lab and not on Mechanical Turk, and thus performed the in-lab version of the experiment.
4.3 Experiment 2 results

As in Experiment 1, we assess whether or not learners formed a productive rule by performing a one-tailed t-test against the hypothesized value of 100%. Participants who formed a productive rule should mark novel nouns with ka on 100% of production test trials, while participants who did not form a rule should do so far less than 100% of the time. Further, we expect participants following the Tolerance Principle to behave categorically, using the regular form 100% of the time if they have formed a productive rule and no more than we would expect by chance if they have not. Thus, in the 3R6E condition, we expect participants to use the regular form no more frequently than any of the other 6 markers (14.29%).

Turning first to our adult control data (shown in Figure 4.2), as in Experiment 1, adults in Experiment 2 do not appear to follow the Tolerance Principle. Adults mark production test trials with the regular form significantly less than 100% of the time in the 5R4E condition \((t(12)=-4.08, p<0.001)\) and significantly more than expected by chance in the 3R6E condition \((t(8)=3.23, p<0.01)\). Instead, adults appear to match the token frequency of the regular form in the language they were exposed to. In Language A, in which the token frequency of ka was 63.27% in the 5R4E condition and 51.02% in the 3R6E condition, adults applied ka to 65.25% of production test trials in the 5R4E condition \((t(7)=0.14, p=0.89)\) and 40.28% in the 3R6E condition \((t(5)=-0.84, p=0.44)\). In Language B, adults again matched the token frequency of the regular form, applying ka not significantly differently from token frequency level in either condition \((5R4E: t(4)=1.08, p=0.34, 3R6E: t(2)=1.70, p=0.23)\).

Turning next to the children, in the 3R6E condition, we find the behavior of children to be well predicted by the Tolerance Principle. That is, children did not use the ka marker significantly differently than one would expect by chance (mean
Figure 4.1: **Usage of plural markers by children in Experiment 2 in the 5R4E and 3R6E conditions.** $R =$ the regular form; $e =$ any exception form; $null =$ unmarked; $other =$ any marker not present in exposure. Error bars are standard error.

= 6.25%, $t(7)=-1.29$, $p=0.89$). However, in the 5R4E condition, we found that the behavior of children was not well predicted by the Tolerance Principle. Children in this condition used the $ka$ marker significantly less than 100% of the time, the value that would indicate productive rule formation ($mean = 42.96$, $t(12)=-4.40$, $p<0.001$), and significantly more than one would expect by chance ($t(12)=-2.21$, $p<0.05$). In other words, when analyzed as the mean usage of the regular form, these data (shown in Figure 4.1) suggest that not only is the usage of $ka$ in the 5R4E condition not consistent with the predictions of the Tolerance Principle, but also children are no longer behaving categorically in this condition. Instead, children use the regular form on an average of 42.96% of production trials.
However, a high standard deviation (46.79%) for the usage of the regular form in the 5R4E condition suggests a more careful look at the individual children in our experiment. When the individual data is inspected (shown in Figure 4.2), it becomes clear that children do, in fact, continue to behave categorically in Experiment 2. Of the 21 children who participated in this experiment, eighteen displayed categorical behavior, using the regular form either 100% of the time (4 children) or no more than expected by chance (14 children).

What could be causing this categorical split in the 5R4E condition? That is, why do some children in the 5R4E condition appear to follow the Tolerance Principle, applying the regular form on 100% of test trials, while other children in the same experimental condition do not. Can we explain why this latter group of children do not form productive rules, even though the Tolerance Principle predicts that they
should? One possibility is that children in Experiment 2 did not learn all of the noun-marker pairings during exposure. We explore this possibility in the following section.

4.4 Calculating the personal Tolerance Principle: revised methods for Experiment 2

In the previous section, we proposed that perhaps children in Experiment 2 were not following the Tolerance Principle because they did not learn all of the noun-marker pairings in our $N=9$ category of nouns. Recall that this value, $N$, is crucial to the calculation of threshold for productivity under the Tolerance Principle ($\theta_N \leq N/\ln(N)$). Thus, whether or not children learned all of the noun-markers pairings during exposure could determine whether or not they should have formed a productive rule. For example, suppose a child in the 5R4E condition, for which productive rule formation is predicted, learned only 7 of the 9 noun-marker pairings. His threshold for productivity under the Tolerance Principle would change from tolerating 4.10 exceptions, calculated for the full category of 9 nouns, to tolerating only 3.60. Further, if 4 of the 7 nouns this child has learned are exceptions, he will have exceeded the number of exceptions a productive rule can tolerate in his category of 7 nouns and is therefore predicted not to form a rule.

To test this hypothesis, we need to reliably assess which of the 9 nouns in our category each child has learned and use this result to calculate a new "personal Tolerance Principle" for each child. Unfortunately, our original rating test (described in Section 4.2.2 was not sufficient to reliably assess this learning. Though our original test sampled all nine nouns in our language, each noun was sampled only once. Therefore, we had no way of determining whether children responded correctly on this single test
item because they had actually learned the noun, or rather because they have simply guessed correctly by chance. Therefore, we revised our rating test part-way through our experiment (7 children had participated in the original version of the rating test). For the revised rating test, we created a two-alternative forced-choice test (2AFC) that sampled each noun four times for a total of 36 rating test trials. On each test trial, one alternative contained the noun paired with the correct plural marker and the other contained the noun paired with an incorrect plural marker. Children were asked to determine which alternative was correct. The 36 test trials were presented in random order.

![Image](image.png)

Figure 4.3: **Example trial for the revised rating test in Experiment 2.** The test is two-alternative forced-choice (2AFC). The correct alternative is shown in (a) and the incorrect alternative is shown in (b) for Language A.

Each 2AFC trial proceeded in the following way. First, a plural image from the exposure was shown flanked by two cartoon children: one dressed in green and the other dressed in purple. Next, the purple cartoon would say one alternative followed by the green cartoon saying the other. Which cartoon spoke first was randomized and which cartoon said the correct answer was counterbalanced across trials. After both alternatives had been presented, children were asked to indicate "which one said the sentence that matched the picture the best?". Children clicked on the cartoon they believed had said the correct sentence.
4.5 Experiment 2 revised results

Of the original 21 children who participated in this experiment, 14 received the revised version of the experiment with this new rating test. Because we are interested in using these methods to classify the categorical nature of children’s behavior in this experiment, the three children who did not behave categorically were excluded from the following analysis.

To assess the number of nouns each child learned, we set a threshold for learning such that a child was considered to "know" a noun if he provided the correct response to 3 or more of the 4 rating test trials for that noun. Next, we counted the total number of nouns each child had learned according to this criteria. This value served as the new $N$ with which to compute each child’s personal Tolerance Principle ($\theta_N \leq N/\ln(N)$). Of these $N$ nouns, we determined how many took the regular form and how many were exceptional ($e$). Finally, we compared the number of exceptions the child had learned during exposure to the child’s new value of $\theta_N$, the number of exceptions the child should tolerate according to their personal Tolerance Principle (see Table 4.2).

To clarify this approach, we illustrate these calculations for a single child in our experiment. Under these criteria, child C22 was found to know 8 noun-marker pairings or $N = 8$. Thus, his personal Tolerance Principle ($\theta_N$) was computed as $8/\ln(8)$, or 3.85. In other words, according to the Tolerance Principle, this child should tolerate 3.85 exceptions before forming a productive rule is no longer the most efficient strategy (not 4.1 as originally calculated for our category of 9 nouns). Of the 8 nouns this child knows, 4 were exceptions, exceeding his threshold of 3.85 exceptions for productivity. Based on his personal Tolerance Principle, this child is predicted not to form a rule.

Note that two children were excluded at this point for failure to learn at least 6 of the 9 nouns during exposure. This threshold was selected because it is not known whether the Tolerance Principle can make accurate predictions for extremely small values of $N$ ($< 6$).
Table 4.2: Calculation and prediction of each child’s personal Tolerance Principle based on the number of noun-marker pairings the child has learned and the number of those that are exceptions.

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>N</th>
<th>e</th>
<th>$\theta_N$</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>C13</td>
<td>6</td>
<td>4</td>
<td>3.35</td>
<td>No Rule</td>
</tr>
<tr>
<td>C14</td>
<td>6</td>
<td>5</td>
<td>3.35</td>
<td>No Rule</td>
</tr>
<tr>
<td>C15</td>
<td>7</td>
<td>4</td>
<td>3.60</td>
<td>No Rule</td>
</tr>
<tr>
<td>C16</td>
<td>7</td>
<td>4</td>
<td>3.60</td>
<td>No Rule</td>
</tr>
<tr>
<td>C17</td>
<td>8</td>
<td>3</td>
<td>3.85</td>
<td>Rule</td>
</tr>
<tr>
<td>C18</td>
<td>7</td>
<td>2</td>
<td>3.60</td>
<td>Rule</td>
</tr>
<tr>
<td>C22</td>
<td>8</td>
<td>4</td>
<td>3.85</td>
<td>No Rule</td>
</tr>
<tr>
<td>C24</td>
<td>7</td>
<td>2</td>
<td>3.60</td>
<td>Rule</td>
</tr>
<tr>
<td>C27</td>
<td>6</td>
<td>2</td>
<td>3.35</td>
<td>Rule</td>
</tr>
</tbody>
</table>

In Figure 4.4, the individual data are re-plotted according to the predictions of the personal Tolerance Principle ("Rule" or "No Rule"), rather than the original 5R4E and 3R6E conditions. The resulting data match very closely with what the personal Tolerance Principle predicts: 8 of the 9 children use the regular form as predicted by their personal Tolerance Principle.

4.6 Experiment 2 Discussion

In Experiment 2, we tested whether the Tolerance Principle would continue to accurately predict productive rule formation when the frequency distribution of the regulars and exceptions was designed more like natural language. While we initially found that children’s behavior was not well predicted by the Tolerance Principle – only children in the 3R6E condition were well predicted by this model – children continued
Recall from Chapter 2 that the Tolerance Principle is intended to apply to the vocabulary of an individual child. That is, the Tolerance Principle is most accurately computed based on the regular and irregular forms that an individual child has acquired. In an effort to explain our categorical results, we modified our experiment such that it was possible to capture an estimate of each child’s individual vocabulary size. We then calculated a personal Tolerance Principle for each child based on the number of regular and irregular forms they had acquired. Our results suggest that the personal Tolerance Principle can more accurately predict the behavior of individual children, suggesting that the Tolerance Principle is an extremely robust metric of productivity.
As in Experiment 1, adults in Experiment 2 continued to match the token frequency of the regular form. This result provides further evidence that the Tolerance Principle may be a mechanism that is exclusive to children.
5.1 Introduction to Experiment 3

In the previous experiments, we have shown that the Tolerance Principle accurately predicts when children will and will not form a productive rule in an artificial language. In Experiment 1, we found that children obey this principle categorically, and confirmed that children follow type and not token frequency during productive rule acquisition. In Experiment 2, we found that child continue to behave categorically, even when the regular form is not longer applied to the most frequent nouns; and that the Tolerance Principle can be used to predict productive rule formation given a child’s individual vocabulary size. In experiment 3, we apply an even more rigorous test to the Tolerance Principle, comparing its predictions for productive rule formation with another approach to quantifying productive rules to determine whether it is, indeed, the Tolerance Principle children are following.

Recall from Section 4.1 that, as the value of $N$ increases, the Tolerance Principle requires an increasingly large majority of items to take the regular from. However, one may ask whether it is really necessary for such a substantial majority to follow the regular pattern. Perhaps learners could follow a much simpler evaluation metric, requiring just a simple majority of forms to follow the regular pattern in order to form a productive rule (e.g., $\geq 50\%$). Table 5.1 shows the threshold for productivity
Table 5.1: Threshold of productivity the Tolerance Principle and Majority of Forms (MoF) evaluation metrics. (Adapted from Yang, 2016).

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\theta_N$</th>
<th>MoF</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>50</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>100</td>
<td>23</td>
<td>49</td>
</tr>
<tr>
<td>200</td>
<td>38</td>
<td>99</td>
</tr>
<tr>
<td>500</td>
<td>80</td>
<td>249</td>
</tr>
<tr>
<td>1000</td>
<td>145</td>
<td>449</td>
</tr>
<tr>
<td>5000</td>
<td>587</td>
<td>2499</td>
</tr>
</tbody>
</table>

For increasing values of $N$ for both the Tolerance Principle and a Majority of Forms evaluation metric. This table illustrates that, in contrast to the Tolerance Principle, as the value of $N$ increases, the number of exceptions a productive rule can tolerate for the Majority of Forms metric continues to increase in proportion to the number of items in the category. Though Yang (2016) finds this method unsatisfactory for to explain which rules are productive and which are restricted in natural language, the results we have presented here are actually consistent with both evaluation metrics. That is, children in our artificial language experiments could have employed formed productive rules based on the Tolerance Principle, or they could have employed the much simpler Majority of Forms evaluation metric. Indeed, for a rule to be productive in a category of 9 nouns, both the Tolerance Principle and the Majority of Forms metric require 5 regular forms. Children in our experiments could have formed a rule in the 5R4E condition but not in the 3R6E condition without the Tolerance Principle ever being invoked.
To address this confound, in Experiment 3, we further modify our artificial language in order to directly test whether children are using the Tolerance Principle or a Majority of Forms evaluation metric to form productive rules. To do so, we increased the number of nouns in our artificial language to 16 such that the Tolerance Principle and Majority of Forms evaluation metrics would predict different thresholds for productivity. In a category of 16 nouns, children following a Majority of Forms evaluation metric should form a productive rule if 9 or more take the regular form. In other words, they should tolerate 7 exceptions. On the other hand, if children are following the Tolerance Principle, their threshold for productivity would be 5.77 exceptions. Thus, they should not form a productive rule unless 11 or more of the noun types take the regular form. We repeat our previous experiment here with this modified language.

5.2 Experiment 3 materials and method

5.2.1 Participants

Ten Children (mean age = 5.97 years, standard deviation = 0.87 years, range = 5.06 – 7.71 years) and seven adult controls (mean age = 20.40, standard deviation = 0.99 years, range = 19.05 – 21.86 years) participated in this experiment. An additional two children and five adults participated but were excluded from analysis for failure to meet experiment inclusion criteria (all 7 had participated in another, similar experiment). Children were recruited from schools, camps, and day-cares in the Washington D.C. metro area and participated in the experiment either at these off-site locations or in our lab. Adult participants were recruited from the Georgetown University community and were run in our lab.
Table 5.2: The number of regular forms required for productive rule formation for the Tolerance Principle and the Majority of Forms evaluation metrics. Shown are the values of $N$ for which the two metrics predict different numbers of regular forms required for productivity as well as margin-of-difference between these predictions.

<table>
<thead>
<tr>
<th>$N$</th>
<th>Tolerance Principle</th>
<th>Majority of Forms</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>14</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>17</td>
<td>13</td>
<td>4</td>
</tr>
</tbody>
</table>

All participants were native English speakers with normal hearing and normal to corrected-to-normal vision. Child participants received stickers and a set of small toys for their participation. Parents of children who participated in our lab received a $10$ travel reimbursement as additional compensation. Adult participants received compensation at a rate of $10$/hour.

5.2.2 Stimuli

To create the artificial language for this experiment, we first computed small values of $N$ for which the Tolerance Principle and the Majority of Forms evaluation metrics require a different number of nouns to take the regular form for productive rule formation (see Table 5.2). Of these values, we elected to create a category of 16 nouns for this experiment. We determined that this number of nouns is both small enough to be learned by children in a single day and allows for a small margin for error, as the number of regular forms the two metrics require differ by 2 (11 regulars for the
Tolerance Principle and 9 regulars for the Majority of Forms). Next, we created a single language condition in which 10 of the nouns took the regular form and 6 of the nouns were exceptions (10R6E). Under these circumstances, the Tolerance Principle and the Majority of Forms evaluation metrics are pitted against each other. If learners obey the Tolerance principle, the threshold for productivity (11 regulars) is not met, and they should not form a productive rule. If on the other hand, learners obey the Majority of Forms metric, they should form a productive rule, as the threshold for productivity is exceeded in this case (9 regulars).

The rest of the language was designed exactly as in Experiment 1 (see Section 3.2.2). As in Experiment 1, we applied the regular plural marker to the most frequent nouns in the Zipfian distribution, such that the token frequency of the regular form was once again extremely high. Further, as in Experiment 1, we created two counterbalanced languages, A and B, such that the rank of the nouns in the Zipfian distribution for language B was the reverse of that of A. For example, in Language A, *mawg* was the most frequent noun and *frag* was least frequent noun, while in Language B, *frag* was most frequent and *mawg* was least frequent. This ensured that there was nothing idiosyncratic about the particular noun frequency ranking used in a given language.

Due to the increased number of nouns in Experiment 3, the exposure corpus generated was 96 sentences (30 singular, 66 plural) rather than the 72 generated for Experiment 1. The production test remained unchanged from Experiments 1 and 2, testing children two times each on a total of six novel nouns (*bleggin, daffin, norg, sep, fluggit, and geed*). The rating test was created as in the revised rating test for Experiment 2 (see Section 4.4), in which children were asked about each noun four times in a 2AFC test for a total of 64 rating test trials. This allows us to determine
Table 5.3: Frequency of nouns and their plural markers in Experiment 3, Language A.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Frequency</th>
<th>N plural</th>
<th>Noun</th>
<th>1OR6E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27</td>
<td>18</td>
<td>mawg</td>
<td>ka</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>10</td>
<td>tomber</td>
<td>ka</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6</td>
<td>glim</td>
<td>ka</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>6</td>
<td>zup</td>
<td>ka</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>4</td>
<td>spad</td>
<td>ka</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>4</td>
<td>daygin</td>
<td>ka</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3</td>
<td>flairb</td>
<td>ka</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>3</td>
<td>kladam</td>
<td>ka</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>lepal</td>
<td>ka</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>3</td>
<td>nerk</td>
<td>ka</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>3</td>
<td>fogul</td>
<td>po</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>3</td>
<td>rov</td>
<td>lee</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>3</td>
<td>nagid</td>
<td>bae</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>3</td>
<td>shen</td>
<td>tay</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>3</td>
<td>mernat</td>
<td>muy</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>3</td>
<td>frag</td>
<td>wio</td>
</tr>
</tbody>
</table>

whether children are capable of learning all of the noun-marker pairings in this much larger 16 noun language.

5.2.3 Procedure

The exposure, production test, and rating test were performed as described in Experiment 1 (see section 3.2.3) with two exceptions. First, adults in Experiment 3 were run in the lab and not on Mechanical Turk, and thus performed the in-lab version of the experiment. Second, both children and adults received the revised 2AFC rating test of Experiment 2 (see Section 4.4).
Figure 5.1: **Usage of plural markers by children and adults in Experiment 3, in which there are 10 regulars and 6 exceptions.**  
$R =$ the regular form; $e =$ any exception form; $null =$ unmarked; $other =$ any marker not present in exposure. Error bars are standard error.

### 5.3 Experiment 3 Results and Discussion

As in Experiments 1 and 2, we assess whether or not learners have formed a productive rule by performing one-tailed t-tests against the hypothesized values of 100% and 14.29% (chance level). Participants who have formed a productive rule should apply the regular form on 100% of test trials, while participants who have not should do so no more than expected by chance. That is, no more than any other marker they were exposed to. If participants are following the Tolerance Principle, they should *not* form a productive rule in this 10R6E condition, applying the regular form no more than expected by chance. However, if participants are following the Majority of Forms evaluation metric, they should form a productive rule, applying the regular form to 100% of production test trials.
Figure 5.1 shows the usage of plural markers by children and adults in this regular, 6 exception condition (10R6E). Focusing first on the child production data, children mark novel nouns with the ka inflection on 39.92% of test trials. This value is significantly lower than 100%, suggesting that children do not follow the Majority of Forms evaluation metric ($t(9)=-4.28, p<0.01$). Further, this value is not significantly different from 14.29%, the level of usage predicted if children follow the Tolerance Principle ($t(9)=1.83, p=0.10$). Taken together, these results suggest that children follow the Tolerance Principle and not a Majority of Forms evaluation metric when forming productive rules.

Turning next to adult production data, we find a familiar pattern of results. As in Experiments 1 and 2, adult usage of the regular form is not significantly different from the token frequency of the regular form ($t(6)=2.04, p=0.09$). However, unlike in Experiments 1 and 2, when we compare adult usage of the regular form to the threshold for productivity (100%), we find that these values are also not significantly different ($t(6)=-1.58, p=0.08$). Thus, our adult results are consistent with either matching the token frequency or following a Majority of Forms evaluation metric. However, recall that a large number of nouns take the regular form in this artificial language and that the regular form is applied to the most frequent nouns in the Zipfian distribution. Combined, these factors lead to an unusually high token frequency: 76.92% of plural tokens included the ka marker. As a result, this feature of our experimental design makes it difficult to determine precisely which metric adults are relying on.

To inspect the data more carefully, we have plotted the usage of the regular form by individual children and adults in Figure 5.2. Visualizing the data in this way allows us to clarify two important points in our analysis. First, as in Experiment 2, there was an extremely high standard deviation in the usage of the regular form for
Figure 5.2: **Usage of the regular form (R) by individual children and adults in Experiment 3, in which there are 10 regulars and 6 exceptions.** Error bars are standard error.

children in Experiment 3 (44.37%). Figure 5.2 allows us to interpret these results by observing that five of the children are, in fact, following the Tolerance principle and not forming a productive rule. However, there are a few sources of variance here. Two children produced the regular form on 100% of trials, consistent with a Majority of Forms interpretation. A further three others behaved non-categorically, producing the regular form probability of the time.

To attempt to resolve this variability across children, we turned again to the results of our 2AFC rating test. We hoped to compute a personal Tolerance Principle for each child as we had in Experiment 2 (see Section 4.4). Perhaps the two children who produced the regular form on 100% of test trials had a different threshold for productivity than the 5 who never produced the regular form. However, upon analyzing the data, we found that five of the ten children did not achieve above chance performance on
the rating test (overall percent correct \(<=50.00\%\)) and four of the children showed signs of test fatigue, quantified as selecting the same value on all but one of the final 20 test trials (e.g., selecting all 2). Only one child fell in both of these groups. Thus 8 of the 10 children were candidates for exclusion on the rating test. We suspect this failure to perform adequately was due to the significantly increased length of this test. Recall that we sample each of the 16 nouns 4 times, resulting in a total of 64 rating test trials. Coupled with the increased exposure length to accommodate the additional noun training (96 exposure sentences instead of the 72 in Experiments 1 and 2), it is possible that children were simply fatigued by the time they completed all 64 rating test trials. Thus, analysis of the rating test is not possible for Experiment 3. Moving forward, we plan to revise this rating test such that children are capable of performing well and collect data from which to interpret our results.

Visualizing the individual data also allows us to clarify our interpretation of whether adults are matching the token frequency or form a productive rule. Of the seven adult participants, four used the regular for 100% on 100% of test trials. This suggests that perhaps adults in this experiment are doing something different than they were in Experiments 1 and 2, in which they were shown to unequivocally match the token frequency. However, with so few subjects, it is difficult to make any definitive claims, and more subjects will be required to interpret these data appropriately. Interestingly, the literature on inconsistent input has shown that when an artificial language is composed of a large number of lexical items, adults tend to regularize the majority token in their input (Hudson Kam & Newport, 2009). It is possible that we are observing a similar effect in this Experiment, in which participants are required to learn a much larger language (16 nouns instead of 9). We explore this possibility with a series of experiments in Chapter 7.
Chapter 6

Is the Tolerance Principle exclusive to children?

6.1 Introduction

While our primary goal in the previous experiments was to ask whether the Tolerance Principle accurately predicts when children will and will not form productive rules, a second but equally important finding emerged. Though the behavior of children is well predicted by the Tolerance Principle in our experiments, the behavior of adults is not. Instead, adult production data closely matches the token frequency of the regular form in the input, indicating that adults may engage a learning mechanism that makes use of this statistic instead. These results suggest that the Tolerance Principle may be a model of productive rule formation that is exclusive to children.

This finding is especially important in the broader context of language acquisition. While it has long been known that there are constraints on the acquisition of language and that these constraints may be maturational in nature, the precise differences between children and adults responsible for childhood being the optimal maturational state in which to acquire language are not yet known. Though a number of hypotheses regarding these differences have been considered, few formal models have been proposed to address why children appear to be optimally suited for the acquisition of language.

In the following sections, we will review the literature on the maturational constraints on language acquisition and discuss the hypotheses that have been proposed
so far. Next, we argue that the Tolerance Principle may be exclusive to children and provide an account of specific differences between children and adults that could explain why children are in an optimal maturational state in which to employ this evaluation metric.

6.2 Maturational constraints on language acquisition

In the language acquisition literature, it has long been proposed that the language input does not provide sufficient evidence with which to acquire the underlying grammar, and the acquisition of language must therefore be constrained in some way (Gold, 1967; Chomsky, 1965; Wexler & Culicover, 1980; Borer & Wexler, 1987; Pinker, 1989). Lenneberg (1967) proposed that constraints on the acquisition of language are maturational in nature, suggesting that language is constrained by a critical or sensitive period in much the same way that other biological systems are. Beyond this period, Lenneberg (1967) proposed that language acquisition would be more difficult, perhaps relying on different underlying learning mechanisms.

Since Lenneberg, a number of studies have provided support for such maturational constraints on language (see Newport, 2002, for review). For example, feral children raised in isolation and not exposed to language before puberty do not go on to achieve native proficiency (Curtiss, 1977). Other populations, whose language proficiency is not confounded by the social and emotional delays resulting from isolation, have also been investigated. Deaf children born to hearing parents, for example, can vary widely in the age at which they are first exposed to sign language\textsuperscript{1}, though they are otherwise typically developing, healthy, and loved. These studies find that

\textsuperscript{1}The majority of deaf children are born to hearing parents who do not know American Sign Language and are often discouraged from learning it. These children are typically first exposed to ASL at school, primarily from children of deaf parents. Further, the age at which parents send their children to schools for the deaf can vary widely.
the ultimate proficiency achieved in a language is highly correlated with the age at which the learner began to acquire the language (Newport, 1990; Mayberry & Eichen, 1991). This same result has been observed for the acquisition of a second language as well (Johnson & Newport, 1989; Long, 1990; Krashen et al., 1979). Johnson & Newport (1989) studied the acquisition of English as a second language in Chinese and Korean adults who had immigrated to the United States at various ages. Age of acquisition was highly correlated with ultimate proficiency in the language, with the earliest exposure resulting in the highest proficiency, even when the number of years of language experience was controlled for.

Importantly, the critical period for language is not intended to reflect an abrupt point beyond which no language learning is possible. Rather, it is more accurately described as a sensitive or optimal period for language acquisition, after which learning is still possible, but is sub-optimal, such that native proficiency may not be achieved. To this point, Johnson & Newport (1989) found that variability in ultimate proficiency was also correlated with age. Those arriving very early in childhood uniformly achieved native proficiency and while variation in individual proficiency increased with age of arrival. This explains why one can occasionally find adult learners who have acquired a second language to native proficiency (e.g., Birdsong, 1992; Bialystok & Hakuta, 1999).

While there is a large body of evidence that maturational constraints on language acquisition exist, it is not yet known precisely what differences between adults and children account for these constraints. One possibility, originally proposed by Lenneberg (1967), suggests that differences in the lateralization of function in the brain can account for these critical period effects. While not directly in support of the hypothesis, several studies indicate that the aforementioned age-related differences in language proficiency can also be demonstrated in the brain. When a second
language is acquired early in childhood, the same brain regions and patterns of activation are observed for both the native language and the second language (Perani et al., 1998). However, if the second language is not acquired until after puberty, the brain regions involved in the first and second language overlap to a much lesser degree, suggesting that the second language has been acquired or is processed differently than the native language (Kim et al., 1997; Perani et al., 1996; Weber-Fox & Neville, 1996; Neville et al., 1997). Further, the degree of overlap is highly variable between individuals after puberty, in line with the variability observed for ultimate proficiency in behavioral investigations (Johnson & Newport, 1989).

Beyond Lenneberg's lateralization hypothesis, one of the most well known proposals to date is Newport's (1990) Less is More hypothesis. This hypothesis suggests that children are optimally suited for language precisely because they are more cognitively limited than adults. That is, the particular cognitive limitations observed in children – a function of the maturational state of childhood – are advantageous for the acquisition of language. Newport (1990) proposes that children’s cognitive limitations, such as memory or processing constraints, force children to "start small." In other words, these limitations constrain children to focus on the smaller, more fundamental components of language during learning, rather than operating over the full complexity of language.

Since Less is More was originally proposed, a number of studies have sought to confirm this hypothesis by demonstrating that starting small is advantageous for language acquisition or by demonstrating that imposing artificial cognitive limitations on adults (e.g., cognitive load) results in better performance on learning tasks\(^2\). For example, computational demonstrations of Less is More have shown that learning models acquire an artificial grammar only when forced to "start small", by initially

\(^2\)See Rohde & Plaut (2003) for review and critique of these tasks
limiting the corpus the model is trained on (Elman, 1991; Goldowsky & Newport, 1993) or otherwise interfering with the model’s memory (Elman, 1993). Similarly, others have demonstrated that imposing artificial cognitive limitations on adults, such as by increasing cognitive load during a task, results in better learning performance and greater ultimate proficiency (Kareev et al., 1997; Cochran et al., 1999; Kersten & Earles, 2001). Still, exactly what these cognitive limitations are and precisely how they manifest as a language learning advantage has not been fully worked out.

6.3 The Tolerance Principle as a model of acquisition

In the following section, I propose that the Tolerance Principle is a model of language acquisition that is unique to children. I argue that specific cognitive differences between children and adults, reminiscent of "starting small" and the memory and cognitive constraints proposed by Newport (1990), result in a maturational state during childhood which is ideally suited for the Tolerance Principle. And, as such, this unique maturational state explains why children engage in the Tolerance Principle but adults do not.

6.3.1 Smaller is better

As reviewed in the previous section, the Less is More hypothesis proposes that one advantage children have over adults is that they "start small", operating over smaller and more fundamental units of language (Newport, 1990). Interestingly, for the Tolerance Principle, the concept of starting small results in a quantifiable computational advantage (see Yang, 2016); that is, the proportion of exceptions a productive rule can tolerate dramatically decreases as the number of items of which the rule may apply ($N$), increases. Figure 6.1 demonstrates that, as $N$ gets very large, a much smaller
Table 6.1: Tolerance Principle threshold of productivity for increasing values of $N$ from Yang (2016).

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\theta_N$</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4</td>
<td>40.0</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>35.0</td>
</tr>
<tr>
<td>50</td>
<td>13</td>
<td>26.0</td>
</tr>
<tr>
<td>100</td>
<td>23</td>
<td>23.0</td>
</tr>
<tr>
<td>200</td>
<td>38</td>
<td>19.0</td>
</tr>
<tr>
<td>500</td>
<td>80</td>
<td>16.0</td>
</tr>
<tr>
<td>1000</td>
<td>145</td>
<td>14.5</td>
</tr>
<tr>
<td>5000</td>
<td>587</td>
<td>11.7</td>
</tr>
</tbody>
</table>

The proportion of exceptions can be tolerated according to the Tolerance Principle. Yang (2016) suggests that this implies "smaller is better" for the acquisition of productive rules. Thus, children and their inherently smaller vocabularies will have an easier time acquiring productive rules than adults. While this likely contributes to the language acquisition advantage enjoyed by children, our results suggest that this cannot be the whole story. In the experiments presented in this dissertation, both children and adults arguably "started small", as the artificial languages they were exposed to contained only a small number of lexical items. Yet, only children follow the Tolerance Principle in our experiments. This suggests that there are additional sources of difference between children and adults that could explain why the Tolerance Principle is unique to children.
6.3.2 Maternal constraints on the storage of regulars and irregulars

Another possibility stems from Newport’s (1990) suggestion that children’s cognitive limitations, particularly memory limitations, are advantageous for language acquisition. Regarding memory, recall that when a productive rule is formed, the Tolerance Principle assumes that the learner has stored the irregulars in a frequency ranked list and that this list must be searched in serial order before the productive rule can be applied. As reviewed in Chapter 2, there is a large body of evidence that supports the plausibility of this assumption. Most notably, there are frequency effects for items hypothesized to be searched in serial order (irregulars/exceptions), but not for those that are composed from a rule (regulars) (Beck, 1997; Prasada et al., 1990; Seidenberg & Bruck, 1990; Ullman, 1999, 2001a).

The aforementioned studies were primarily conducted with adults, which might lead one to argue that both adults and children share this memory organization. However, adults in these experiments were tested on their native language, which they are known to have acquired during the critical period for language acquisition. Thus, these rules and exceptions are likely organized according to the systems in place during that childhood maturation state. Further, one can directly address whether this is the case by investigating whether these same frequency effects appear for a second language learned after puberty. As suspected, in contrast to the results observed for learner’s native language, frequency effects are observed for both regular and irregular forms in a second language (Babcock et al., 2008; Brovetto & Ullman, 2001; Neubauer & Clahsen, 2009). This suggests that second language learners store both regular and irregular forms as if they were exceptions and thus never form a
productive rule. Or, in terms of the Elsewhere Condition, a productive rule is never applied because a more specific form exists.

One prominent hypothesis as to the neural mechanism underlying this storage of regulars and exceptions suggests that they are handled by two different underlying memory systems: procedural and declarative memory (see Ullman, 2001b, for review). The declarative memory system (Mishkin et al., 1984; Squire & Zola, 1996; Schacter & Tulving, 1994), thought to be responsible for memories for facts and events, is subserved by a hippocampal network primarily including the medial temporal lobe and its connections to temporal cortex (Suzuki & Amaral, 1994). According to Ullman (2001b, page 718), this system may be "particularly important for the learning of arbitrarily related items – that is, for the associative/contextual binding of information". On the other hand, the procedural memory system (Mishkin et al., 1984; Squire & Zola, 1996; Schacter & Tulving, 1994) is thought to be responsible for the learning of patterns and new skills and is subserved primarily by the basal ganglia and its connections to the frontal cortex, particularly the left inferior frontal gyrus (IFG) (Mishkin et al., 1984; Squire & Zola, 1996; Schacter & Tulving, 1994). This system may be important "in grammatical-structure building – that is, in the sequential and hierarchical combination of stored forms ('walk' + 'ed') and abstract representations" (Ullman, 2001b, page 718).

Thus, according to the declarative/procedural model, the irregular forms are predicted to be stored in declarative memory while the regular form is predicted to be composed in real-time by the procedural memory system (Ullman et al., 1997; Ullman, 2001b). A number of sources of evidence contribute support for this hypothesis. In general, these studies find that damage to the basal ganglia and/or the left IFG results in impairment for the regular form (Marin et al., 1976; Badecker & Caramazza, 1987; Coslett, 1986; Gopnik, 1994; Ullman et al., 1997; Ullman, 1998a, 1999; Van der Lely
& Ullman, 2001; Ullman et al., 2005), while damage to the hippocampal network results in impairment for irregular forms (Clahsen & Almazan, 1998; Bromberg et al., 1994; Ullman et al., 1997; Cappa & Ullman, 1998; Patterson et al., 2001; Ullman et al., 2005). For example, individuals with a particular hereditary form of specific-language impairment (SLI) have been shown to have abnormalities in the left IFG and basal ganglia (Vargha-Khadem et al., 1998). When these individuals were assessed on English past tense production, as predicted, they were unable to apply the productive "add -d" rule to novel verbs (Clahsen et al., 1997; Gopnik, 1994; Ullman & Gopnik, 1999). Further, frequency effects for both regular and irregular forms were observed, suggesting that both regular and irregular forms were memorized (Clahsen et al., 1997; Gopnik, 1994; Ullman & Gopnik, 1999). On the other hand, individuals with Alzheimer’s disease tend to have damage localized to the temporal lobe (Boller & Duyckaerts, 1997) and, as such, show more difficulty producing irregular than regular forms (Cappa & Ullman, 1998). Additionally, patients with more severe impairments in object naming and fact retrieval make more errors producing the irregular form (Ullman et al., 1997; Ullman, 1998b).

What does this declarative/procedural model have to do with memory differences between children and adults? It turns out that these two memory systems may undergo substantial changes over the course of development. It has been hypothesized that the procedural memory system is functioning at its peak during childhood, progressively declining from adolescence through early adulthood (Fredriksson et al., 2000; Schlaug, 2001; Walton et al., 1992; Wolansky et al., 1999; Janacek et al., 2012), while the declarative memory shows the opposite developmental pattern (Campbell & Spear, 1972; DiGiulio et al., 1994; Kail & Hagen, 1977; Meudell, 1983; Ornstein, 1978; Siegler, 1978; Finn et al., 2016). In support of this hypothesis, recent research investigating gray matter volume changes across development has found that the volume
of basal ganglia structures decreases with age while the volume of the hippocampus shows an inverted u-shaped trajectory (Wierenga et al., 2014) (see Figure 6.1). While this hypothesis does not translate exactly as limitations to memory as suggested in the Less is More hypothesis (Newport, 1990), it does suggest that children are in a particular maturational state with respect to these memory systems and perhaps that this maturational state is optimized for the acquisition of productive rules. That is, children have enhanced procedural memory, thought to be responsible for the composition of rules, and lower declarative memory function, thought to be responsible for the storage of irregular forms. Thus, perhaps the Tolerance Principle is capturing something fundamental about productive rule formation under the constraints of this particular maturational state in development.

![Figure 6.1](image)

Figure 6.1: Gray matter volume for structures underlying the declarative and procedural memory circuits across development (ages 7 to 24 years), adapted from Wierenga et al. (2014). The caudate and putamen (basal ganglia), structures underlying the procedural memory system are shown in (a) and the hippocampus, a structure underlying the declarative memory system is shown in (b).

### 6.3.3 Probability matching and cognitive control

Another way in which children may be cognitively limited is in the development of the pre-frontal cortex (PFC). Indeed, humans have been shown to undergo a long
period of prolonged prefrontal immaturity\(^3\) (Rakic et al., 1986; Chugani & Phelps, 1986; Huttenlocher & Dabholkar, 1997). As a result, human children are impaired at a number of abilities known to rely on mature PFC function (Diamond & Doar, 1989). In particular, children are known to have impaired cognitive control abilities compared with adults.

There are a number of circumstances in which adult-like cognitive control abilities are clearly advantageous. For example, in the Stroop task, participants must name the ink color that a color word is printed in (e.g., the word "red" printed in green ink) (Stroop, 1935). Performance requires participants to engage their cognitive control abilities to suppress the meaning of the word in order to produce the correct ink color. Adults are significantly better at this task than children (MacLeod, 1991; Demetriou et al., 2001; Adleman et al., 2002; Hanauer & Brooks, 2003).

However, some have hypothesized that these same cognitive control abilities may result in a disadvantage during certain kinds of learning (e.g., Thompson-Schill et al., 2009). That is, during learning processes that depend on low-level competition mechanisms, cognitive control abilities may actually interfere. A common example comes from the literature on probability learning (e.g., Neimark, 1956; Gardner, 1957, 1958; Weir, 1972). To illustrate, imagine you are presented with two light bulbs and are told to guess which one will light up. You notice that the light bulb on the left lights up 75% of the time, while the light bulb on the right lights up 25% of the time, so you guess the light bulb on the left 75% of the time. This behavior is known as *probability matching* and is commonly observed in adults performing such tasks (Gardner, 1957; Weir, 1964, 1972). However, to achieve the maximum number correct, the optimal strategy is actually to always guess the left light bulb. This strategy is called *max-\(^3\)In other primate species, the development of the brain occurs at roughly the same rate across all cortical areas (Rakic et al., 1986)
imization, and it is well known that young children employ this strategy in such tasks (Stevenson & Weir, 1959; Weir, 1964; Derks & Paclisanu, 1967). Why do adults adopt a sub-optimal strategy in this task? One possible answer has been proposed by Thompson-Schill et al. (2009, page 3):

"Why do you use a less optimal decision strategy than your toddler? One possibility is that your well-developed PFC-mediated cognitive control system allows you to override brute-strength competition and guess: In an unregulated competition between alternate responses, the most frequent form dominates (i.e., maximization). In order to make a less frequent (but potentially goal-relevant) response (i.e., probability matching), a control mechanism intervenes. You do badly because you can guess, unlike your toddler who cannot."

Thus, one possibility is that the Tolerance Principle is capturing some sort of low-level competition mechanism that is optimized for acquiring rules. Adults in our experiments, with their fully developed cognitive control abilities, are able to override this low-level competition mechanism and make (what they believe to be) an educated guess: I heard the "ka" inflection 65% of the time, so I'll guess "ka" 65% of the time. Indeed, as mentioned in our previous experiments, this probability matching behavior has been observed in similar artificial language learning experiments (Hudson Kam & Newport, 2005, 2009). Children on the other hand, with their limited cognitive capacity (Newport, 1990), are unable to override the low-level competition mechanism, and are thus more optimally suited for the acquisition of productive rules.
In the previous sections, we have argued that the Tolerance Principle may be a model of productive rule formation that is exclusive to children and proposed a number of hypotheses as to why this may be the case. Rather than depending on any one of the hypotheses alone, it is more likely some combination of all three (and perhaps others we have not yet proposed) that creates an optimal maturational state in which to acquire a productive rule. After this period, adults may begin to store everything as an exception in declarative memory and/or override low-level selection mechanisms using their mature cognitive control systems, resulting in a failure to acquire productive rules.

In the following chapters, we begin to test these hypotheses by attempting to make adults behave like children. That is, we attempt to impose cognitive limitations on adults that are analogous to those we have seen in children. Under these circumstances, will adults follow the Tolerance Principle in the same way that children do? We explore the answers to these questions in the following chapter.
Chapter 7

Experimental approaches to adults and the Tolerance Principle

7.1 Introduction

In the previous chapter, we argued that the Tolerance Principle may be a model of productive rule formation that is exclusive to children. Following Newport (1990) and the Less is More hypothesis, we proposed that children have particular cognitive limitations that, when combined, create an optimal state in which to acquire productive rules. In the present chapter, we begin to address this hypothesis in a series of experiments in which we attempt to make adults behave more like children. We do so by creating artificial cognitive limitations in adult learners to determine whether or not they will follow the Tolerance Principle under these circumstances.

In Experiment 4, we begin by asking whether adults will follow the Tolerance Principle if the language they are acquiring contains a significantly larger number of lexical items. In previous experiments, such an approach has been found to increase cognitive demands on adult learners, therefore leading them to acquire artificial languages in a manner more similar to children (Hudson Kam & Newport, 2009). In Experiments 5, we further limit our adult participants by imposing a time pressure on the production phase of our task. This effectively impairs their ability to rely on cognitive control abilities and forces them to depend on more low-level competition mechanisms. In both cases, we hypothesize that adults will begin to behave more like children when these artificial cognitive limitations have been imposed.
7.2 Adults acquiring a larger artificial language

The primary goal of Experiment 4 was to impose cognitive limitations on adult learners by exposing them to a much larger artificial language. As mentioned above, previous experiments have found that adults under these circumstances begin to move away from frequency matching behavior and toward behavior that is more consistent with what is observed in children (Hudson Kam & Newport, 2009). To test whether a larger language would similarly push adults to engage the Tolerance Principle, rather than token frequency matching, we exposed adults in Experiment 4 to an artificial language that contained 36 nouns instead of only 9.

7.2.1 Experiment 4 materials and method

Participants

83 Adults participated in this experiment. Of these 83, 14 participated in the High-Frequency Regular version, 33 participated in the Matched Token Frequency version, and 36 participated in the Farther From Threshold version (described below). All adults were native English speakers with normal to corrected-to-normal hearing and vision. Participants were recruited and run on Mechanical Turk and received compensation at a rate of $10/hour.

Stimuli

The goal of this experiment was to test whether adults would follow the Tolerance Principle if the language they were acquiring contained significantly more lexical items. To do so, we created an artificial language with 36-nouns (four times as many as in Experiment 1). We then computed the Tolerance Principle threshold for this value of \( N \) to determine the number of exceptions a productive rule could tolerate.
in a language of this size: \( \theta_N \leq 10.05 \). Thus, we can test the Tolerance Principle by creating two conditions: one in which the Tolerance Principle predicts a productive rule should be formed (fewer than 10.05 exceptions) and one in which a productive rule should not be formed (greater than 10.05 exceptions).

Importantly, in order to thoroughly test whether there were any circumstances under which adults would follow the Tolerance Principle in this 36-noun language, we designed three different versions: High-Frequency Regular, Matched Token Frequency, and Farther From Threshold.

The High-Frequency Regular and Matched Token Frequency versions were similar in that the conditions selected to test the Tolerance Principle were the same for both versions. That is, participants were predicted to form a productive rule if they were exposed to 26 regulars and 10 exceptions (26R10), but not if they were exposed to 24 regulars and 12 exceptions (24R12E). However, they differed in the frequency of the regulars and exceptions. In the High-Frequency Regular version, the regulars were applied to the highest frequency nouns in the Zipfian distribution (as in Experiment 1), such that the token frequency was high in both conditions. In the Matched Token Frequency version, the exceptions were more evenly distributed with the regulars in the top of the Zipfian distribution (as in Experiment 2), such that the token frequency of the regular form was matched in both conditions.

The Farther From Threshold version was created to ensure that we were not inadvertently setting the Tolerance Principle up for failure. In a 36-noun language, the Tolerance Principle predicts that 10.05 exceptions should be tolerated, and in the previous conditions, participants are predicted to form a productive rule if there are 10 exceptions, but not if there are 12. We reasoned that in this much larger language, it may be necessary to allow the Tolerance Principle a bit more margin for error. To achieve this, in the Father From Threshold version, participants were predicted to
form a productive rule if they were exposed to 32 regular and 4 exceptions (32R4E), but not if they were exposed to 18 regulars and 18 exceptions (18R18E). In this version, the regular form was again applied to the highest frequency nouns (As in our Experiment 1 and the High-Frequency Regular version of this experiment). The nouns and distribution of plural markers used in all three experiment versions are shown in Table 7.1.

Table 7.1: Frequency of nouns and their plural markers in Experiment 4 for the High-Frequency Regular (HFR), Matched Token Frequency (MTF), and Farther From Threshold (FFT) conditions.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Freq</th>
<th>Noun</th>
<th>HFR</th>
<th>MTF</th>
<th>FFT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>26R</td>
<td>24R</td>
<td>32R</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>flerbit</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>mang</td>
<td>ka</td>
<td>ka</td>
<td>po</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>flobut</td>
<td>ka</td>
<td>ka</td>
<td>meg</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>melnog</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>mazner</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>nerk</td>
<td>ka</td>
<td>ka</td>
<td>tay</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>ladnuh</td>
<td>ka</td>
<td>ka</td>
<td>tay</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>mernot</td>
<td>ka</td>
<td>kum</td>
<td>kum</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>blergenfol</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>dugolu</td>
<td>ka</td>
<td>li</td>
<td>li</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>dilbu</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>flugerdo</td>
<td>ka</td>
<td>ler</td>
<td>ler</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>bampogin</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>blagu</td>
<td>ka</td>
<td>suh</td>
<td>suh</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>lomba</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>gerko</td>
<td>ka</td>
<td>bon</td>
<td>bon</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>sulto</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>zamper</td>
<td>ka</td>
<td>gi</td>
<td>gi</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>pernisel</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>kowoltu</td>
<td>ka</td>
<td>ka</td>
<td>fu</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>gentu</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>klamin</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>miktlu</td>
<td>ka</td>
<td>ka</td>
<td>ka</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>blifen</td>
<td>ka</td>
<td>ka</td>
<td>mib</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>mawg</td>
<td>ka</td>
<td>po</td>
<td>ka</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>misnu</td>
<td>ka</td>
<td>meg</td>
<td>ka</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>ferluku</td>
<td>po</td>
<td>tay</td>
<td>ka</td>
</tr>
</tbody>
</table>

95
Continuation of Table 7.1

<table>
<thead>
<tr>
<th>Rank</th>
<th>Freq</th>
<th>Noun</th>
<th>HFR</th>
<th>MTF</th>
<th>FFT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>26R</td>
<td>24R</td>
<td>26R</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td>rungnot</td>
<td>meg</td>
<td>kum</td>
<td>ka</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>rinola</td>
<td>tay</td>
<td>li</td>
<td>ka</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>slergen</td>
<td>kum</td>
<td>ler</td>
<td>ka</td>
</tr>
<tr>
<td>31</td>
<td>2</td>
<td>fumpogu</td>
<td>li</td>
<td>suh</td>
<td>ka</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>nagru</td>
<td>ler</td>
<td>bon</td>
<td>ka</td>
</tr>
<tr>
<td>33</td>
<td>2</td>
<td>blerfe</td>
<td>suh</td>
<td>gi</td>
<td>ka</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>kerno</td>
<td>bon</td>
<td>bip</td>
<td>ka</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>melamu</td>
<td>gi</td>
<td>fu</td>
<td>ka</td>
</tr>
<tr>
<td>36</td>
<td>2</td>
<td>fogul</td>
<td>bip</td>
<td>mib</td>
<td>ka</td>
</tr>
</tbody>
</table>

End of Table

**Procedure**

The exposure, production test, and rating test were performed as described in Experiment 1 (see section 3.2.3) with only slight modifications to the production and rating tests, described below.

The production test was the same as in all previous experiments, except that each novel noun was paired with a different novel image (shown in Appendix A, Table A.3). As in the previous experiments, participants were tested on each of the six novel nouns (bleggin, daffin, norg, sep, fluggit and geed) twice.

The rating test was the same as in Experiment 1, in which we sampled a total of four nouns from the exposure set: two that followed the regular form and two that were exceptions in both conditions. In Experiment 4, we again selected nouns such that two were High-Frequency and two were low frequency, choosing rank 1, 2, 8, and 9 for this experiment. While 8 and 9 may seem highly ranked in a list of 36 nouns, the Zipfian distribution is such that the majority of nouns are in the long tail of low frequency. Indeed, all nouns below frequency rank 7 are attested only 3 times each in the exposure corpus, once in singular form and twice in plural form (see Table 7.1).
Table 7.2: Token frequencies of the regular form in Experiment 4.

<table>
<thead>
<tr>
<th>Experiment Version</th>
<th>Condition</th>
<th>Token Fq.</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Frequency Regular</td>
<td>24R10E</td>
<td>83.33</td>
<td>79.33</td>
<td>20.69</td>
<td>7.82</td>
</tr>
<tr>
<td></td>
<td>26R12E</td>
<td>80.00</td>
<td>78.25</td>
<td>19.25</td>
<td>7.28</td>
</tr>
<tr>
<td>Matched Token Frequency</td>
<td>24R10E</td>
<td>68.33</td>
<td>67.99</td>
<td>28.75</td>
<td>7.19</td>
</tr>
<tr>
<td></td>
<td>24R10E</td>
<td>68.33</td>
<td>63.99</td>
<td>29.70</td>
<td>7.20</td>
</tr>
<tr>
<td>Farther From Threshold</td>
<td>32R4E</td>
<td>78.00</td>
<td>63.95</td>
<td>34.27</td>
<td>8.31</td>
</tr>
<tr>
<td></td>
<td>18R18E</td>
<td>55.00</td>
<td>63.81</td>
<td>31.36</td>
<td>7.72</td>
</tr>
</tbody>
</table>

7.2.2 **Experiment 4 results**

As in all previous experiments, we determine whether learners in our experiment are following the Tolerance Principle by comparing their usage of the regular form to the hypothesized values of 100% and the level of chance. That is, learners who have formed a productive rule should apply the regular form to novel nouns 100% of the time, while learners who have not formed a productive rule should apply the regular form no more than is predicted by chance (or in other words, no more than any other plural marker). For the High-Frequency Regular and Matched Token Frequency versions, chance level is 7.70% in the relevant condition (24R12E). In the Farther From Threshold version, chance level is 5.26% in the relevant condition (18R18E).

Further, to determine whether participants match the token frequency of the regular form, rather than following the Tolerance Principle, we compare usage of the regular form to the token frequency of the regular form for each experiment version. Token frequencies for each experiment version are shown in Table 7.2.

In all three experiment versions, adults continued to match the token frequency of the regular form in their input. Participants’ use of the regular form was significantly
Figure 7.1: **Usage of the regular form by adults in each version of Experiment 4.** X-axis corresponds to the conditions in each version for which the Tolerance Principle predicts a productive rule should ("Rule") or should not ("No Rule") be formed. Rule/No Rule corresponds to conditions 26R10E/24R12E in the High-Frequency Regular and Matched Token Frequency versions, and 32R4E/18R18E in the Farther from Threshold version. The black horizontal lines indicate the token frequency of the regular form. Error bars are standard error.
Table 7.3: Summary of t-test results in Experiment 4.

<table>
<thead>
<tr>
<th>Experiment Version</th>
<th>Cond.</th>
<th>Value</th>
<th>Mean</th>
<th>t(df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Frequency Regular</td>
<td>26R10E</td>
<td>100</td>
<td>79.33</td>
<td>t(6) = −2.64</td>
<td>&lt; 0.05 *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>83.33</td>
<td>79.33</td>
<td>t(6) = −0.51</td>
<td>= 0.62 n.s.</td>
</tr>
<tr>
<td></td>
<td>24R12E</td>
<td>7.69</td>
<td>78.24</td>
<td>t(6) = 9.70</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80.00</td>
<td>78.25</td>
<td>t(6) = −0.24</td>
<td>= 0.82 n.s.</td>
</tr>
<tr>
<td>Matched Token Frequency</td>
<td>26R10E</td>
<td>100</td>
<td>67.99</td>
<td>t(15) = −4.45</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.33</td>
<td>67.99</td>
<td>t(15) = −0.05</td>
<td>= 0.96 n.s.</td>
</tr>
<tr>
<td></td>
<td>24R12E</td>
<td>7.69</td>
<td>63.99</td>
<td>t(16) = 7.81</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.33</td>
<td>63.99</td>
<td>t(16) = 0.39</td>
<td>= 0.70 n.s.</td>
</tr>
<tr>
<td>Farther From Threshold</td>
<td>32R4E</td>
<td>100</td>
<td>63.95</td>
<td>t(16) = −4.34</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78.00</td>
<td>63.95</td>
<td>t(16) = −1.69</td>
<td>= 0.11 n.s.</td>
</tr>
<tr>
<td></td>
<td>18R18E</td>
<td>5.26</td>
<td>63.81</td>
<td>t(16) = 7.70</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55.00</td>
<td>63.81</td>
<td>t(16) = 1.16</td>
<td>= 0.26 n.s.</td>
</tr>
</tbody>
</table>

different from what is predicted by the Tolerance Principle in all three versions, and not statistically different from the token frequency of the regular form. A summary of these data is shown in Table 7.3.

7.2.3 EXPERIMENT 4 DISCUSSION

As in Experiments 1–3, adults in Experiment 4 continued to match the token frequency of the regular form, even though the languages they were acquiring contained many more lexical items. These results are not consistent with previous findings that suggest that adult learners will begin to behave more like children when the experimental language they are acquiring contains significantly more lexical items (Hudson Kam & Newport, 2009).

There are a few differences between previous work and the current experiment that may explain why adults continue to match the token frequency of the regular
form here. First, while we increased the size of our noun category substantially (from 9 to 36), consistent with the previous work, our language as a whole was significantly smaller than that used in Hudson Kam & Newport (2009). In our language, all sentences were intransitive, composed of only a single noun phrase (Verb + Noun + Marker). Further, our Verb category contained only a single verb, "gentif". In the Hudson Kam & Newport (2009) work, the language had significantly more complex syntax, including both transitive and intransitive sentences, negation, and a much larger number of verbs. Thus, it is possible that our language was not sufficiently complex to drive adults away from their token frequency matching strategy.

Second, the previous experiments exposed adults to input that was inconsistent – that is, every form was probabilistically marked (e.g., "mawg" is marked with ka 67% of the time and po 33% of the time). In the present experiment, every form was consistently marked, such that a given type would always take a particular marker (e.g., "mawg" is always marked with ka). Perhaps there are some underlying differences in the acquisition of these two types of language stimuli that result in language size effecting one but not the other.

7.3 Adults performing the wug test under time pressure

In the previous experiment, we found that increasing the size of our artificial language was not sufficient to drive adults away from their token frequency matching strategy. Thus, in Experiment 5, we take another approach to limiting the cognitive capacity of adults in an experimental setting.

In Chapter 6, we proposed that one possible way in which children were more cognitively limited than adults was in cognitive control. We argued that perhaps adults use their fully developed cognitive control abilities to override lower-level selection
mechanisms in order to generate an educated guess about which inflection to provide in our experiments. This suggests that if one could limit the cognitive control abilities of adults, they might resort back to the underlying mechanisms whose behavior may be better predicted by the Tolerance Principle.

To test this hypothesis, we replicated our initial experiment (Experiment 1) in adults under circumstances in which their cognitive control abilities were impaired. To do so, we asked adults to complete the production test under a time pressure, such that they would be unable to execute cognitive control mechanisms to decide which inflection they should mark novel forms with. Under these circumstances, we hypothesize that adults should rely on the Tolerance Principle to form a productive rule.

7.3.1 Experiment 5 Materials and Method

Participants

Thirty-one adults (mean age = 19.88 years, standard deviation = 2.01 years, age range = 18 - 26 years) participated in this experiment. An additional two adults began the experiment but failed to complete it due to equipment malfunction. All participants were native English speakers with normal hearing and normal to corrected-to-normal vision recruited from the Georgetown University community. The experiment was conducted in our lab and participants were compensated at a rate of $10 per hour with a $5 bonus incentive, described below.

Stimuli

In Experiment 5, participants were exposed to the same 9 noun artificial language as in Experiment 1 (described in Section 3.2.2). Recall that the Tolerance Principle
Figure 7.2: Example trial for the timed production test in Experiment 5. Participants were required to provide the plural sentence within the 1.5 second response window.

The threshold for a category of 9 nouns is $\theta_N \leq 4.10$, and participants are therefore predicted to form a productive rule if there are fewer than 4.10 exceptions, but not if there are more. Thus, as in Experiment 1, we created two language conditions: one where a productive rule should be formed (5R4E: 5 regulars, 4 exceptions) and one where a productive rule should not be formed (3R6E: 3 regulars, 6 exceptions). As in all previous experiments, nouns followed an approximately Zipfian distribution and we counterbalanced the noun frequency ranking by creating two languages: A and B. Language A and B differed only in that, in language A, *mawg* was most frequent and *lepal* was least frequent, while in language B, this noun ranking was reversed. Refer to table 3.1 in Chapter 3 for the frequency of nouns and their plural markers for conditions 5R4E and 3R6E in Language A.

**PROCEDURE**

The exposure, production test, and rating test were performed as described in Experiment 1 (see Section 3.2.3) with three exceptions. First, participants in Experiment 5
were run in the lab and not on Mechanical Turk, and thus performed the in-lab version of the experiment. Second, all participants received the revised 2AFC rating test as in Experiment 2 4.4, in which participants were asked about each noun 4 times. Third, and most important to our experiment, participants were given a 1.5 second time limit with which to complete each trial on the production test. To increase the salience of the time limit, immediately following the presentation of the singular form, a green screen appeared and participants heard a beep to indicate the start of the trial. The plural test image appeared immediately after and participants were given a 1.5 second response window in which to produce the plural sentence. When the time limit expired, the screen turned red and participants heard a beep indicating the end of the trial. A sample production test trial is illustrated in Figure 7.2. To encourage participants to produce sentences as quickly as possible, they were told that they would receive a small bonus for each trial they were able to finish within the time limit, up to an extra $5.

7.3.2 Experiment 5 results

To begin, we asked whether adults match the token frequency of the regular for in the input, as they had in all previous experiments. In both the 5R5E and 3R6E conditions, participants applied the regular form significantly more than the token frequency of the regular form (5R4E: t(12)=10.98, p<0.001; 3R6E: t(12)=4.52, p<0.001). Thus adults appear to have moved away from the matching the token frequency of the regular form under these circumstances.

---

1The 1.5 second response window was selected after some experiment piloting determined that 1.5 seconds was just long enough for most participants to produce a plural sentence (e.g., "gentif mawg ka") in the language.

2At the end of the experiment, all participants received the full $5 bonus regardless of their performance on the production test.
Table 7.4: Summary of t-test results in Experiment 5.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Value</th>
<th>Mean</th>
<th>t(df)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5R4E</td>
<td>100%</td>
<td>97.38%</td>
<td>t(12) = -1.32</td>
<td>=0.17 n.s.</td>
</tr>
<tr>
<td></td>
<td>75.51%</td>
<td>97.38%</td>
<td>t(12) = 10.98</td>
<td>&lt; 0.001***</td>
</tr>
<tr>
<td>3R6E</td>
<td>100%</td>
<td>84.94%</td>
<td>t(12) = -2.64</td>
<td>&lt; 0.05 *</td>
</tr>
<tr>
<td></td>
<td>59.18%</td>
<td>84.94%</td>
<td>t(12) = 4.52</td>
<td>&lt; 0.001***</td>
</tr>
<tr>
<td></td>
<td>14.29%</td>
<td>84.94%</td>
<td>t(12) = 12.39</td>
<td>&lt; 0.001***</td>
</tr>
</tbody>
</table>

Are adults following the Tolerance Principle under these circumstances, then? In the 5R4E condition, we find that adults produce the regular form on 97.38% of production test trials, not significantly different from the hypothesized value of 100% (t(12) = -1.32, p = 0.17). This suggests that adults in the 5R4E condition do, indeed, obey the Tolerance Principle under these circumstances. In the 3R6E condition, however, we find that adults apply the regular form on 84.94% of trials. This is significantly higher than what we would expect by chance (t(12) = 12.39, p < 0.001), suggesting that adults do not obey the Tolerance Principle in this language condition.

7.3.3 Experiment 5 discussion

In Experiment 5, we placed adults under a time pressure during our production test to determine whether imposing a limitation on cognitive control would lead adults to rely on the Tolerance Principle. We find that these circumstances do, indeed, push adults away from the matching the token frequency of the regular form in their input. Neither adults in the 5R4E condition nor those in 3R6E condition matched the token frequency of the regular form. Instead, we found that adults in the 5R4E condition
appeared to match the Tolerance Principle, while adults in the 3R6E condition did not.

When we inspect the individual data (Figure 7.3, we can see that a substantial number of adults applied the regular form in both the 5R4E and 3R6E conditions under these circumstances. Given that the Tolerance Principle predicts a categorical distinction between rules that are productive (5R4E) and rules that are not (3R6E), these results lead us to suspect that adults may be relying on an underlying mechanism other than the Tolerance Principle under these circumstances. One possibility is that the Tolerance Principle is not an appropriate model for adult behavior due to additional differences between adults and children (perhaps the differences in the storage of regular and irregular forms suggested in section 6.3.2). Thus, while our time
pressure prevented adults from engaging in probability matching, further differences between adults and children exist. Based on the data in figure 7.3, adults may rely on a simple frequency based mechanism under these circumstances, in which they regularize the majority token present in their input.

7.4 Discussion

The primary goal of Chapter 7 was to begin to test our hypotheses that children’s cognitive limitations lead them to be optimally suited for the acquisition of productive rules as predicted by the Tolerance Principle. We hypothesized that imposing cognitive limitations on adults, either by increasing the size of the language (Experiment 4) or imposing a time pressure on our production test (Experiment 5), would lead them to follow the Tolerance Principle in our experiments. Interestingly, we found that adults do not follow the Tolerance Principle under either of these circumstances. When we increased the size of the language (Experiment 4), adults continued to match the token frequency as they had in all previous experiments. When we imposed a time constraint on the production test (Experiment 5), adults moved away from this frequency matching behavior. However, they did not move toward following the Tolerance Principle. Instead, our results suggest that adults regularize the majority token when their cognitive faculties are limited in this way.

As mentioned in Chapter 6, there is not likely a single difference between children and adults to which our results can be attributed. Rather, there is more likely a complex interaction between a number of cognitive differences that create the optimal circumstances for the learning mechanism that are well predicted by Tolerance Principle.
Chapter 8

Conclusion

8.1 Summary of the dissertation

The primary goal of this dissertation was to use artificial language learning experiments to test whether the Tolerance Principle (Yang, 2005, 2016) accurately predicts the formation of productive rules during language acquisition. In Experiment 1, we exposed learners to a 9-noun artificial language and found that children formed a productive rule when there were 5 regulars and 4 exceptions, but not when there were 3 regulars and 6 exceptions; just as the Tolerance Principle predicts. In Experiment 2, we found that children continue to follow the Tolerance Principle when the underlying frequency distribution of the regular and exceptional forms is structured more like that of natural language. Further, that the Tolerance Principle can predict productive rule formation based on an individual child’s vocabulary size. In Experiment 3, we compared the Tolerance Principle to a simple majority of types evaluation metric and found that children follow the Tolerance Principle and not the majority of types under these circumstances.

Across Experiment 1–3, while our results suggest that children follow the Tolerance Principle in the acquisition of productive rules, adults in these experiments do not. Rather, adults appear to match the token frequency of the regular form in their exposure. This observation led us to argue that Tolerance Principle may be model of productive rule formation that is special for children. We argued that particular
cognitive limitations during children, in the spirit of Newport’s (1990) Less is More hypothesis, create the optimal conditions in which to acquire productive rules, which are well predicted by the Tolerance Principle model. In Experiments 4 and 5, we begin to address this hypothesis by imposing cognitive limitations on adults to determine whether they will rely on the Tolerance Principle under those circumstances. While learners in Experiment 5 did move away from matching the token frequency of the regular form, neither experimental manipulation resulted in adults following the Tolerance Principle. These results suggest that there are a number of differences between children and adults that act in concert to create the optimal circumstances for productive rule acquisition.

8.2 Implications for acquisition of productive rules

As reviewed in Chapter 2, the Tolerance Principle is an evaluation metric (Chomsky, 1965) which allows both children and scientists to determine the productive status of rules in natural language. Together with the previous corpus analyses (Yang, 2016), the experiments in this dissertation make a convincing case that the Tolerance Principle is the most accurate model of productive rule acquisition to date. However, it is important to point out that whether or not the Tolerance Principle is exactly right is not necessarily its most interesting aspect. Rather, the most interesting and useful thing about the Tolerance Principle is that it offers a testable prediction to which we can compare empirical data.

As such, we still have plenty of work ahead of us. We can ask how the Tolerance Principle fares against other metrics of productively, and whether it predicts children’s behavior in artificial languages that require nested generalizations along specific features (e.g., German plural). Eventually, if we push hard enough, we may
find the place at which the Tolerance Principle fails to accurately predict behavior in children. Far from disappointing, such a finding would allow us to apply what we have learned about productivity and children’s behavior to revise and improve the Tolerance Principle, at which point we would start the whole process over again. This same sentiment is true for any theory or model in language acquisition. The utility of such models is not whether they are exactly correct, but that we can use them to inform what we know about how acquisition proceeds.

8.3 Implications for investigating maturational constraints in language acquisition

As reviewed in Chapter 6, researchers have long known that there are constraints on the acquisition of language and that at least some of these constraints are maturational in nature. However, the precise nature of the differences between children and adults that lead childhood to be the optimal time at which to acquire language is not yet known. As such, discovering the nature of these constraints remains one of the central problems in the field of language acquisition.

A powerful method for investigating possible mechanisms and constraints involves combining models like the Tolerance Principle with empirical experiments that compare children and adults, as demonstrated in this dissertation. Such combinations allow us to test precise predictions about the nature and developmental trajectory of aspects of language acquisition. Further, we can design experiments with which to explicitly test hypotheses about the underlying differences between adults and children, as demonstrated in Chapter 7.
The work in this dissertation highlights the progress that can be made in language acquisition by drawing on strengths from both linguistics and psycholinguistics. Indeed, there is a rich literature in linguistics on the acquisition of language, from which we can derive powerful theories and models of acquisition. Then, using empirical approaches from psycholinguistics, we can test how accurately these models predict the behavior of actual children acquiring natural or artificial languages. Without these empirical tests, it is difficult to be sure that children indeed acquire language as linguistic theories predict. Similarly, without these theories and models of acquisition, psycholinguists have no predictions to experimentally test.

Importantly, empirical tests investigating the nature of language acquisition need to be conducted on child subjects. As evidenced by this dissertation and other literature in the field (e.g., Berko, 1958; Singleton & Newport, 2004; Hudson Kam & Newport, 2005; Austin, 2010), children often behave in a way that is quite different from adults on language acquisition tasks. Though this fact has long been known, recent trends in the study of language acquisition suggest that this point needs reminding. In recent years, the focus of many acquisition studies has shifted to computationally demanding probabilistic models, big data, and large-scale studies conducted with adult participants online. As such, it is my hope that the results in this dissertation serve as a reminder that these kinds of investigations can only take us so far. In order to investigate how children acquire natural language, we need to conduct experiments with child learners. When we do develop models of acquisition, it is important to consider the cognitive resources children are likely to have available to them, and to rigorously test these models against the behavior of actual children. In
other words, a statistical description of the data is not sufficient; a specific prediction that can be empirically evaluated in children in the lab is required.

8.5 Future directions

Finally, there are a number of different directions in which to pursue this work in the future. First, as mentioned above, one could evaluate the Tolerance Principle against other evaluation metrics, or ask whether children can acquire nested productive rules based on particular sub-classes of features (e.g., German plural). Second, further experiments investigating differences between adults and children, with more targeted approaches to imposing cognitive limitations, could contribute to our understanding of the underlying maturational differences.

One interesting possibility often considered of language acquisition mechanisms is to ask whether the Tolerance Principle is specific to productive generalization in language or whether it is modeling some domain general mechanism for learning. Yang (2016) suggests that the Tolerance Principle may be applied to other types of productive generalizations beyond language and offers a slight reformulation of the Tolerance Principle – the Sufficiency Principle – that quantifies whether a child has experience sufficient evidence to make a generalization. Future work will ask whether this is indeed the case, and if so, whether this kind of domain-general learning mechanism is unique to children as well.
Appendix A

Experimental Stimuli

A.1 Exposure Sets

The exposure nouns and corresponding image stimuli for the artificial languages used in Experiments 1–5 are shown in Tables A.1 and A.2. The 9 noun language was used in Experiments 1, 2, and 5; the 16 noun language was used in Experiment 3; and the 36 noun language was used in Experiment 4.

A.2 Production Test Set

The six novel nouns used in the production test are shown in Table A.3 with their corresponding image stimuli. The same production test was used in all five experiments.

A.3 Image Origins

The images used as experimental stimuli are called *glitch cubimals* and are part of the art assets for an online game called Glitch by Tiny Speck. After the game shut down in 2012, its entire library of art and code was released into the public domain under the Creative Commons CC0 1.0 Universal License. All assets are freely available for unrestricted use and can be downloaded from [https://www.glitchthegame.com/public-domain-game-art/](https://www.glitchthegame.com/public-domain-game-art/).
Table A.1: **Experimental stimuli for the 9 and 16 noun artificial languages.** The 9 noun language was used in Experiments 1, 2, and 5. The 16 noun language was used in Experiment 3 and included all stimuli from the 9 noun language.

<table>
<thead>
<tr>
<th>9 Noun Language</th>
<th>16 Noun Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>Image</td>
</tr>
<tr>
<td>mawg</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>tomber</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>glim</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>zup</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>spad</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>daygin</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>flairb</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>klidam</td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>lepal</td>
<td><img src="image" alt="Image" /></td>
</tr>
</tbody>
</table>
Table A.2: **Experimental stimuli for the 36 noun artificial languages.** The 36 noun language was used in Experiment 4.

<table>
<thead>
<tr>
<th>Noun</th>
<th>Image</th>
<th>Noun</th>
<th>Image</th>
<th>Noun</th>
<th>Image</th>
<th>Noun</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>flerbit</td>
<td><img src="lerbit.png" alt="Image" /></td>
<td>dugolu</td>
<td><img src="dugolu.png" alt="Image" /></td>
<td>pernisel</td>
<td><img src="pernisel.png" alt="Image" /></td>
<td>rungmot</td>
<td><img src="rungmot.png" alt="Image" /></td>
</tr>
<tr>
<td>mang</td>
<td><img src="mang.png" alt="Image" /></td>
<td>dilbu</td>
<td><img src="dilbu.png" alt="Image" /></td>
<td>kowoltu</td>
<td><img src="kowoltu.png" alt="Image" /></td>
<td>rinola</td>
<td><img src="rinola.png" alt="Image" /></td>
</tr>
<tr>
<td>flobuf</td>
<td><img src="flobuf.png" alt="Image" /></td>
<td>flugerdo</td>
<td><img src="flugerdo.png" alt="Image" /></td>
<td>gentu</td>
<td><img src="gentu.png" alt="Image" /></td>
<td>slergen</td>
<td><img src="slergen.png" alt="Image" /></td>
</tr>
<tr>
<td>melnog</td>
<td><img src="melnog.png" alt="Image" /></td>
<td>bambogin</td>
<td><img src="bambogin.png" alt="Image" /></td>
<td>klamin</td>
<td><img src="klamin.png" alt="Image" /></td>
<td>fumpogu</td>
<td><img src="fumpogu.png" alt="Image" /></td>
</tr>
<tr>
<td>mazner</td>
<td><img src="mazner.png" alt="Image" /></td>
<td>blagur</td>
<td><img src="blagur.png" alt="Image" /></td>
<td>miktu</td>
<td><img src="miktu.png" alt="Image" /></td>
<td>nagru</td>
<td><img src="nagru.png" alt="Image" /></td>
</tr>
<tr>
<td>nerk</td>
<td><img src="nerk.png" alt="Image" /></td>
<td>lomba</td>
<td><img src="lomba.png" alt="Image" /></td>
<td>blifen</td>
<td><img src="blifen.png" alt="Image" /></td>
<td>blerfe</td>
<td><img src="blerfe.png" alt="Image" /></td>
</tr>
<tr>
<td>ladnuh</td>
<td><img src="ladnuh.png" alt="Image" /></td>
<td>gerko</td>
<td><img src="gerko.png" alt="Image" /></td>
<td>mawg</td>
<td><img src="mawg.png" alt="Image" /></td>
<td>kerno</td>
<td><img src="kerno.png" alt="Image" /></td>
</tr>
<tr>
<td>mernot</td>
<td><img src="mernot.png" alt="Image" /></td>
<td>sulto</td>
<td><img src="sulto.png" alt="Image" /></td>
<td>misnu</td>
<td><img src="misnu.png" alt="Image" /></td>
<td>melanu</td>
<td><img src="melanu.png" alt="Image" /></td>
</tr>
<tr>
<td>blergenfol</td>
<td><img src="blergenfol.png" alt="Image" /></td>
<td>zamper</td>
<td><img src="zamper.png" alt="Image" /></td>
<td>ferluku</td>
<td><img src="ferluku.png" alt="Image" /></td>
<td>fogul</td>
<td><img src="fogul.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Table A.3: **Experimental stimuli for the production tests.** The same production test was used in Experiments 1, 2, 3, and 5. A separate production test was used in Experiment 4.

<table>
<thead>
<tr>
<th>Experiments 1–3, 5</th>
<th>Experiment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>Image</td>
</tr>
<tr>
<td>bleggin</td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>daffin</td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>norg</td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>sep</td>
<td><img src="image7.png" alt="Image" /></td>
</tr>
<tr>
<td>flugit</td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td>geed</td>
<td><img src="image11.png" alt="Image" /></td>
</tr>
</tbody>
</table>
The following is a summary of the Tolerance Principle as derived by Yang (2016, pages 60-66). Refer to Yang (2016) for a more detailed explanation.

The Tolerance Principle assumes that the distribution of word frequencies in an given sample obey Zipf’s law (Zipf, 1949). As such, in a given sample of unique word types \( \{w_1, w_2, \ldots, w_N\} \), the rank, \( r_i \), and frequency, \( f_i \), of a word, \( w_i \), are inversely proportional. This is expressed in equation B.1 for some constant \( C \).

\[
r_i f_i = C
\]  

Thus the probability with which \( w_i \) is expected to occur, \( p_i \), can be expressed as:

\[
p_i = f_i \sum_{k=1}^{N} f_k = \frac{C}{r_i} \sum_{k=1}^{N} \frac{C}{r_k} = \frac{1}{i H_N}
\]

where \( H_N = \sum_{k=1}^{N} \frac{1}{k} \) (B.2)

Under the serial search assumption, it takes \( r \) steps to reach the \( r \)th word in a list of \( N \) items. Thus, the time complexity required accessing a word when everything has been stored as a frequency-ranked list, \( T(N, N) \), is:

\[
T(N, N) = \sum_{r=1}^{N} r \frac{1}{r H_N} = \frac{N}{H_N}
\]  

To compute the time complexity required for storing only the exceptions and forming a productive rule, \( T(N, e) \), Yang (2016) assumes that the expected time complexity for accessing the exceptions (\( T(e, e) \)) is \( e/H_e \) as in equation B.3. The
expected time for accessing the remaining items \((N - e)\) is simply the number of exceptions, \(e\). Thus, \(T(N, e)\) is computed as:

\[
T(N, e) = \frac{e}{N} T(e, e) + (1 - \frac{e}{N})e = \frac{e}{N} e + (1 - \frac{e}{N})e \tag{B.4}
\]

To solve the equation \(T(N, N) = T(N, e)\), Yang (2016) first approximates the \(N\)th harmonic number, \(H_N\), with the natural log of \(N\) \((\ln N)\). Thus, the equation can be solved for \(e\) as:

\[
x \frac{e}{\ln e} + (1 - x)e = \frac{N}{\ln N} \tag{B.5}
\]

Dividing both sides by \(N\) results in the remaining derivation (see B.6–B.8).

\[
x^2 \frac{1}{\ln N + \ln x} + (1 - x)x = \frac{1}{\ln N} \tag{B.6}
\]

\[
f(x) = x^2 \frac{1}{\ln N + \ln x} + (1 - x)x - \frac{1}{\ln N} \tag{B.7}
\]

\[
f\left(\frac{1}{\ln N}\right) = \frac{(1/\ln N)^2}{\ln N + \ln\ln N} + (1 - \frac{1}{\ln N}) \frac{1}{\ln N} - \frac{1}{\ln N}
= -\left(\frac{1}{\ln N}\right)^2 + \left(\frac{1}{\ln N}\right)^3 \frac{\ln N}{\ln N + \ln\ln N} \approx -\left(\frac{1}{\ln N}\right)^2
\approx 0 \text{ for large values of } N \tag{B.8}
\]


120


Fleischhauer, E. & Clahsen, H. (2012). Generating inflected word forms in real time: Evaluating the role of age, frequency, and working memory. In Biller, A.,


logical patterns. *Language and cognitive processes, 8*(1), 1–56.

are retrieved from memory but regular forms are rule generated. *Bulletin of the
Psychonomic Society, 28*(6), 519–519.

*Language, 7*41–788.

Rakic, P., Bourgeois, J.-P., Eckenhoff, M. F., Zecevic, N., & Goldman-Rakic, P. S.
(1986). Concurrent overproduction of synapses in diverse regions of the primate

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does
not require a rule. *Cognitive Psychology, 45*(1), 45–94.

city in Optimality Theory*, 1–16.


Rumelhart, D. & McClelland, J. (1986). On learning the past tenses of english
verbs. In *Parallel distributed processing: explorations in the microstructure of cog-


Ullman, M. T. (1998a). Evidence that lexical memory is part of the temporal lobe declarative memory, and that grammatical rules are processed by the frontal/basalganglia procedural system. *Brain and Language.*
Ullman, M. T. (1998b). Evidence that lexical memory is part of the temporal lobe declarative memory, and that grammatical rules are processed by the frontal/basal-ganglia procedural system. *Brain and Language.*


