Genes and Environment in Language Acquisition: A Study of Early Vocabulary and Syntactic Development in Twins

by

Jennifer B. Ganger

B.A., Linguistics
University of Michigan, 1992

Submitted to the Department of Brain and Cognitive Sciences in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Cognitive Science

at the

Massachusetts Institute of Technology

September 1998

Signature of Author:______________________________
Department of Brain and Cognitive Sciences
August 24, 1998

Certified by:______________________________
Steven Pinker
Professor of Cognitive Science
Director, McDonnell-Pew Center for Cognitive Neuroscience
Thesis Co-supervisor

and by:______________________________
Kenneth Wexler
Professor of Psychology and Linguistics
Thesis Co-supervisor

Accepted by: ______________________________
Gerald Schneider
Professor of Neuroscience
Chair, Department Graduate Committee

Genes and Environment in Language Acquisition: A Study of Early Vocabulary and Syntactic Development in Twins

by

Jennifer B. Ganger

Submitted to the Department of Brain and Cognitive Sciences August 24, 1998 in Partial Fulfillment of the Requirements for the degree of Doctor of Philosophy in Cognitive Science

ABSTRACT

The goal of this thesis is to explore the contributions of heredity and environment to language development using the twin method. This method consists of comparing monozygotic (MZ) twins to dizygotic (DZ) twins to see if MZs have a more similar course of development. Two major aspects of language development are examined: vocabulary and syntax.

The development of vocabulary in 43 MZ and 33 same-sex DZ English-speaking twins was studied via parent report. The rate and content of the first 100 or so words in children's productive vocabularies were examined. MZ twins had more similar ages at reaching several milestones of vocabulary growth, more similar overall rates of vocabulary growth, more similar spurts in vocabulary growth, and more similarity in actual words and categories of words they produced than DZ twins, suggesting heritability in early word learning. However, the effects of environment were also prominent in vocabulary development, dwarfing effects of heritability.

The development of first word combinations and of first overregularizations of the past tense rule "add *-ed*" (e.g., *goed* for *went*) were also studied by parent report to examine the contributions of heredity and environment to the development of syntax. Both the age of producing first word combinations and the rate of producing combinations were much more similar in MZ than in DZ twins (N = 24 MZ, 23 DZ). This effect showed large heritability and small effects of environment. The rate of producing past tense overregularizations also appeared to be more similar in MZ than in DZ twins, though this result is tentative due to the small number of twins in this study (16 MZ, 11 DZ).

Finally, the development of verb tense and agreement was studied longitudinally in 4 MZ and 4 DZ twin pairs. The percentage of correct use of tense/agreement morphemes over time was on average more correlated in the MZ than in the DZ twins, suggesting that differences in the development of the Inflectional system also have genetic influence.

Thesis Supervisors:	Steven Pinker	Kenneth Wexler
Titles:	Professor of Cognitive Science; Director, McDonnell-Pew Center for Cognitive Neuroscience	Professor of Psychology and Linguistics

Acknowledgements

The work presented in this thesis is the result of an enormous amount of effort—more work than one person could possibly hope to accomplish in five years. It could not have been completed without the contributions of many, many individuals. Allison Baker, Sonia Chawla, Melanie Soderstrom, and Emily Wallis have all been co-authors on some parts of this work and have been of enormous help in collecting, analyzing, and preparing the data in this thesis. Allison Fitch-Markham, Eva Palmerton, and Aparna Polavarapu also deserve special thanks for their work in the summer of 1998 that went directly into this thesis.

In addition to these people, I have been lucky enough to work with no less than 37 research assistants from MIT, Harvard, Wellesley, and other Universities over the past three years. I would like to acknowledge all of them: Lya Ana Batlle, Stacy Betz, Johanna Bobrow, Maya Cantwell, Todd Chamoy, Marian Chen, Chris Chow, Brandy Evans, Dina Finan, Jennifer Frank, Natalie Garner, Leo Gonzalez, Nina Gray, Nate Greenslit, Areej Hassan, Chris Huffman, Andy Hinkhouse, Ai Kato, Victoria Keith, Anna Khasin, Cathy Lavelle, Cherry Liu, Chris Lim, Rebeccca Lubens, Sue Mosher, Regina Moucka, Tanis O'Connor, Alice Oh, Stefani Okasaki, Eva Palmerton, Sunbin Song, Agi Stachowiak, Gabriel White, Elisabeth Winterkorn, Emily Wishneusky, Josh Wise, and Stephanie Yang. Thanks also to Yaron Koren, who did some programming work.

My committee members, Steve Pinker, Ken Wexler, Liz Spelke, David Pesetsky, and Kim Saudino, have also been of a source of help and inspiration. Steve Pinker and Ken Wexler, my co-advisors and also co-authors on some parts of this thesis, have been extremely generous both intellectually and financially and never wavered in their enthusiasm for this project even when I did. Liz Spelke met with me several times in the spring of 1998 and continues to give insightful and thought-provoking comments. David Pesetsky has been an inspirational teacher and collaborator throughout my time at MIT and I feel very fortunate to have known him. Finally, Kim

Saudino, as the first behavior geneticist to read this work, has pointed out numerous conceptual and technical problems. I would like to thank her for taking the time to serve on my committee.

Several other scientists have been of help to me in completing this project over the last three years, including Karin Stromswold, Jeff Gilger, and Mabel Rice.

I also thank my family and friends. My husband Greg Ganger has been supportive and sympathetic throughout graduate school. My parents, Esther and Larry Melamed, have been supportive throughout my academic career and especially helpful in this difficult summer of 1998. Cristina Sorrentino and Fei Xu have also been good friends and colleagues during this summer and I hope they continue to be.

The work in this thesis was funded by NIH grant #HD18381 to Steven Pinker and an NSF Research Training Grant (Language Acquisition and Computation, #DIR9113607) to the department.

Table of Contents

Chapter 1

Genes and environment in language development: A twin study

Faced with the difficult problem of acquiring language, it is likely that our species has converged on one solution and that qualitative variation in that solution is minimal. Indeed, in many ways, the process of language development is highly uniform. Children typically begin expressive language with babbling, go on to produce single words in isolation, then produce word combinations of increasing length and complexity until, by the age of four, normal children have just about mastered the grammar of their language (e.g., Lenneberg 1967; Morley 1965; Brown 1973; Stromswold 1990). Furthermore, consistent with a universal ability, achieving these milestones is not heavily dependent on experience: Lenneberg, for example, described cases of deaf children who began babbling at the appropriate age and tracheotomized children who very rapidly caught up to their peers in language development after the removal of their breathing tube. Both of these examples demonstrate that very early language unfolds on a maturational timetable that is only mildly affected by experience.

However, despite the uniformity and robustness of the *quality* of language development, researchers long ago noticed that there is substantial *quantitative* variation in the timing of this process (e.g., Nelson 1973; Bates, Bretherton & Snyder 1988). For instance, there are individual differences among children in the rate of word learning (Nelson 1973; Goldfield & Reznick 1990; Bates *et al.* 1988), in the age of mastering inflectional morphemes (Brown 1973; Rice & Wexler 1996), and in the pace of acquiring more subtle syntactic knowledge. Although these differences have not been denied by proponents of the universals paradigm, they have been downplayed because of overwhelming qualitative similarities. In contrast, the present work is based on the premise that there is much to be learned from the etiology of individual differences in language development. By dividing the source of these individual differences into genetic (or *heritable*) and environmental (or *non-heritable*) components, it is possible to establish that there is a genetic influence on variation in language development, thereby supporting the view that language is a genetically encoded skill that unfolds according to a maturational timetable.

The goal of this work is to establish that genes play a role in the development of two major aspects of language: vocabulary and syntax. In the realm of vocabulary, our focus is on the stage in which words are produced in isolation—roughly the first 100 vocabulary items. Individual differences in this "one-word stage" have a long history of study in developmental psychology, but differences among children in both rate and content have usually been attributed to environmental differences, such as number of siblings, styles of parenting, and television watching (e.g., Nelson 1973). However, such studies, which reported correlations between environmental factors and characteristics of language development, failed to investigate possible genetic influences. The current study fills that gap by showing that genetic differences also factor into individual differences in word learning.

In the realm of syntax, three studies have been undertaken. The first concerns multiple-word combinations: those earliest of sentences, often consisting of just two words, that children typically begin producing between 18 and 24 months of age. This "two-word stage" has also been the subject of much scrutiny by psychologists (e.g., Brown & Fraser 1963; Bloom, Lightbown & Hood 1976; Braine 1976; Nelson 1981), who have sometimes argued that individual differences in the use of these phrases reflect fundamentally different ways of learning language. Genetic influences on the age of onset of these proto-sentences have never before been investigated, though. This study may reveal that individual differences in the earliest signs of productive syntax have a genetic basis.

The second syntax project concerns a well-studied test case of a grammatical rule, the English past tense rule "add *–ed*" (e.g., Marcus *et al.* 1992). By studying overregularizations (e.g., *I falled*), we can track the onset and development of this rule in twins. This rule serves as a case study for grammatical rules in general, and heritability of this rule would suggest that syntax more generally in heritable. Taken together, the goal of these two studies (the word combinations and overregularization studies) is to show that individual differences in the development of syntax have a genetic basis.

Finally, the third study of syntax concerns the development of verbal inflection (tense and agreement, or *Infl*). In this smaller and more intensive study, we seek to show that inflection in particular has a genetic basis, above and beyond whatever genetic effects are found in syntax generally. The theory underlying this study is Wexler's Optional Infinitives model (Wexler 1994).

By showing that there is genetic influence on individual variation in language development, we can argue that genes must be influential in language development. This is our primary goal. A secondary goal is to estimate the heritability and the environmentality of variance vocabulary and syntax and to compare them to each other. If their heritabilities (and concomitant environmental contributions) are different, that difference would suggest that vocabulary and syntax are learned by different mechanisms and may represent fundamentally different processes. The method used to both establish genetic effects and estimate heritability is the twin method, first proposed by Galton (1876) and now an established method of behavior genetics.

1.1 Using individual differences to measure heritability: The twin method

The twin method allows us to divide sources individual variation into heritable and non-heritable factors and to estimate the relative magnitude of those factors . To understand how this works, it is first necessary to consider population or phenotypic variation (V_P) in a trait as being composed of genetic variance and environmental variance, as expressed in Eq. 1. (For the moment, the multiple kinds of genetic and environmental variance, as well as interactions and correlations between genetic and environmental factors, are suppressed. These will be considered below.)

Eq 1. $$V_P = V_G + V_E$$

V_g is the component of phenotypic variance (V_p) that is due to genetic variance, and the proportion of V_P that is made up of V_G (i.e., $\frac{V_G}{V_P}$) is known as *heritability* (h^2)[1]. Genetic variance is the result of the differential distribution of alleles among individuals. It constitutes the variation that natural selection feeds on and theoretically eliminates in the course of evolution. However, it is possible for variant alleles to co-exist for a variety of reasons. For instance, they may not result in differential adaptiveness, they may be the result of recent mutations, or the very potential for heterozygosity might confer some selective advantage (as in the case of an allele for one of the genes that determines the shape of the hemoglobin molecule in humans: in the homozygote, this allele causes sickle-cell anemia, but in the heterozygote it confers some resistance to malaria Griffiths *et al.* 1993).

Genetic variance can be divided into two main sources: additive and non-additive. Additive genetic affects are expressed to the same extent in every individual in which they are present and they combine with other genes in proportion to their contribution. In other words, an individual with two different alleles at a genetic locus (a heterozygote) will show expression of that pair of genes in a manner that is intermediate between individuals who are homozygous for one version of the gene and individuals who are homozygous for the other. Non-additive genes do not have this property, but instead have one of two types of relationships with other genes. The first type of non-additive genetic effect is the dominant/recessive one. In a dominance relationship, the two alleles of a gene at a single locus interact with each other such that the dominant allele suppresses the recessive allele's expression; the two alleles do not blend in an additive fashion. The second non-additive genetic effect is epistasis, in which the same sort of relationship exists between genes but at two or more different loci. The effects of non-additive genetic variance on the estimation of heritability will be considered below.

[1] Technically, the statistic described in the main text is known as *broad sense* heritability, because it is based on both additive and non-additive genetic factors (see below in main text). *Narrow sense* heritability (which would be impossible to calculate in the present study but is sometimes possible when the breeding history of an animal is known extensively) reflects only additive genetic factors. However, I will refer to broad sense heritability as heritability (h^2) throughout.

The other source of variation is V_E: variance in the environment or any non-inherited source. The environment can take many forms, from simple events such as injuries or diseases to complex, multi-faceted influences like upbringing and culture. Environmental effects can be divided into two sources: shared and non-shared. Shared environmental variance is variance in the environment that is shared by co-twins or any family members and that serves to make them more similar. Artifacts like systematic rater bias can also inflate this term (Neale & Maes in prep; Neale & Stevenson 1989). Non-shared environmental variance is the variance among individuals that arises from unique experiences of any kind and that decreases the similarity of twins or other family members. Random error also inflates this quantity. The proportion of shared environmental variance over the total phenotypic variance is known as c^2, while the proportion of non-shared environmental variance is known as e^2.

In addition to these two main sources of variance— V_G and V_E —two additional sources of variance are relevant: the correlation between V_G and V_E (gene-environment correlation), and the interaction of V_G and V_E (gene-environment interaction, V_G x V_E).

Gene-environment correlation can occur when individuals affect their own environment because of their genotype—for instance, a child's genotype may affect the way his parents treat him, so that his genes and his environment are not independent. Gene-environment correlations can also occur because most individuals have their environment provided for them by family members, who have similar genotypes. Thus when children are influenced by environments provided by their parents' genes, they are also being influenced by their own genes, because about half of them also exist in their parents.

Gene-environment interaction occurs when the very same genotype is expressed differently in different environments or the same environment acts differently on different genotypes. A simple example of this effect is that individuals who are genetically susceptible to a disease may or may not express it depending on environmental pathogens, while those who are not susceptible will not express the disease no matter what the environment. A more subtle example of this effect is a finding by Heath & Martin (1986), cited in Neale & Maes (in prep), that the heritability of depression in female twins is higher if they are single than if they are married, suggesting that an environmental factor—marriage—has an effect on the expression of genes for affect. Other examples of the gene-environment interaction can be found in animal and plant studies (e.g., Mather & Jinks 1982; see also Wahlsten 1994 for a review). For now I will make the risky but quite conventional move (e.g., Plomin, DeFries & McClearn 1990) of ignoring any correlation or interaction between genes and environment as a separate sources of variance, but will return to the consequences of this decision below.

Heritability and its counterparts, shared and non-shared environment, may be estimated in a variety of ways. The one used in this study is the twin method (Galton 1876), which is based on the comparison of identical and fraternal twins. An idealized description of the twin study, which will be re-examined in section 1.4, is as follows. Identical or *monozygotic* (MZ) twins result from the division of a single zygote or fertilized ovum. This split results in two embryos, and ultimately two human beings, who share 100% of their genes. Ideally, we could simply compare traits in MZ twins to calculate genetic variance, since the extent to which MZ twins are similar reflects their shared genetic variance. However, most MZ twins are raised in the same home by the same parents, resulting in similar (shared) environments, so it would never be clear how much of their similarity resulted from their shared environment in addition to their shared genes.[2] In other words, the correlation for any trait between MZ co-twins is the sum of their genetic variance, of which they share 100%, and their (shared) environmental variance, of which they also have 100% in common, divided by the total variance. These quantities are readily converted to h^2 and c^2. These relationships are expressed in Equation 2.

Eq. 2 $$r_{MZ} = \frac{V_G + V_E(shared)}{V_P} = h^2 + c^2$$

---

[2]One way to tease genetic and environmental variance apart in MZ twins is to study twins raised in different environments—so-called cases of "separation at birth." Such cases have been studied extensively by Thomas Bouchard (e.g., Bouchard, Lykken, McGue, Segal, & Tellegen 1990) but are not be used in the present study as they are extremely difficult to obtain.

From Equation 2 it should be clear that it is impossible to estimate h^2 or c^2 if only the MZ correlation is known. Fortunately, nature has provided a second equation: the correlation between fraternal or dizygotic (DZ) twins. DZ twins result from the fertilization of two ova from the same mother by two sperm from the same father.[3] These twins share on average 50% of their segregating genes, just like any pair of full siblings.[4] Crucially, though, they have environments that are as similar as those for MZ twins, since they are normally raised in the same home by the same parents. The correlation for any trait between fraternal co-twins is made up the proportion of their shared phenotype provided by their shared environmental variance (which has a coefficient of 1.0) and that provided by their shared genetic variance. V_G has a coefficient of 1/2 for DZ twins, since they have on average only half their genes in common.

Eq. 3 $$r_{DZ} = \frac{\frac{1}{2}V_G + V_E(shared)}{V_P} = \tfrac{1}{2}h^2 + c^2$$

The proportion of genetic variance, or heritability, can now be found by subtracting the DZ correlation from the MZ correlation and multiplying the result by 2. Note that c^2 drops out, so our idealization that co-twins share 100% of their environments is unimportant, as long as both MZ and DZ twins have equally similar environments.

$$r_{MZ} - r_{DZ} = \frac{V_G + V_E - \left(\frac{1}{2}V_G + V_E\right)}{V_P} = h^2 + c^2 - \tfrac{1}{2}h^2 - c^2$$

$$r_{MZ} - r_{DZ} = \frac{\frac{1}{2}V_G}{V_P} = \tfrac{1}{2}h^2$$

Eq. 4 $$h^2 = 2\left(r_{MZ} - r_{DZ}\right)$$

---

[3]There have also been reported cases of dizygotic twins resulting from the fertilization of two ova from the same mother by sperm from two different fathers. These twins are perhaps more aptly dubbed "half-fraternal" since they are as related to each other genetically as half-siblings. Furthermore, with the advent of *in-vitro* fertilization, other dizygotic twin types may arise.

[4]It has been suggested that DZ twins share more than 50% of their genome—that is, that they are genetically more similar than ordinary siblings—as evidenced by lower rates of skin graft rejection (Geschwind 1983). If that is the case, then the twin method as it stands will underestimate heritability, since it assumes that DZ twins share only 50% of their genetic variance. However, the skin graft effect could be due the fact that DZ twins gestate together and are breastfed during the same time period after birth, perhaps giving them similar immunities.

Once heritability is calculated in the manner described, the components of environmental variance can be computed. Shared environment, c^2, is computed by subtracting the heritability from the MZ correlation ($r_{MZ} - h^2$; see Eq 2). Non-shared environment, e^2, is calculated indirectly as one minus the MZ correlation ($1 - r_{MZ}$).

1.2 The DeFries-Fulker method and the estimation of heritability

Although it is possible to compute heritability in exactly the manner described above, by first calculating MZ and DZ correlations, DeFries & Fulker (1985, 1988) proposed a simpler method that does all the calculations in one step and can easily provide confidence intervals and significance testing for h^2 with standard statistical software. Although their method was developed for proband selected data (i.e., data for which one or both members of a twin pair is selected for deviance on some measure), it works with non-selected data as well (Plomin, DeFries & McClearn 1990).

The method consists of regressing twins' scores (T) on their co-twins' scores (C) and including two other independent variables. The first is a term that estimates the genetic relationship (R) between each pair of twins: 1.0 for all MZs and 0.5 for all DZs. The second is the interaction of the co-twins' scores and the relationship term (C x R). Explicitly, the model is

$$T = B_1C + B_2R + B_3C \times R$$

Each pair of twins is entered twice, once with a given twin as T and his/her co-twin as C, and again with the assignments reversed. The reason for such "double-entry" is that, unless one twin is a proband, there is no principled reason for making one twin T and the other C, and the choice affects their correlation. To solve this problem, each pair is entered twice, and concomitant corrections are made in the standard error (and resulting statistics) in the output of statistical programs to correct for the artificially doubled sample size.[5]

---

[5]Corrected standard error (SE) (from Stevenson *et al.* 1993) is:

$$\text{Corrected SE} = \text{Obtained SE} \times \sqrt{(N_d - k - 1)/(N_s - k - 1)}$$

where

B_1 is a measure of shared environmental variance (c^2), because it is the co-twin resemblance with their genetic relationship partialed out. B_2 estimates the extent to which the mean for probands (in proband selected data) has a genetic etiology and so does not concern the present research. B_3 estimates heritability: a significant regression coefficient for this term indicates that a model in which MZ twins are twice a correlated as DZ twins does a good job of describing the data. Mathematically, B_3 yields the same estimate of heritability as that described in the last section. Finally e^2 may be calculated as $1 - h^2 - c^2$.

This regression method, the DeFries-Fulker or "DF" method, will be used throughout this work to estimate heritability. To simply establish genetic influence, which is our first goal, it is only necessary to show that MZ twins are significantly more similar than DZ twins using any test. Thus in many cases in the work to follow, instead of using correlations or the DF method to estimate heritability, t-tests are used simply to establish genetic effects. These tests are typically performed on difference scores between MZ twin pairs versus difference scores for DZ pairs, so that we expect MZ difference scores to be smaller than DZ differences scores if MZ twins are more similar. Note that this trend is the reverse of what we expect to find if we check correlations: MZ correlations should be larger than DZ correlations if variance in a trait has genetic influence.

1.3 Some comments on heritability and the twin method

It should be clear from the discussion above that heritability is a population statistic and is dependent on the particular population variances on which it is based. If these variances differ in different populations, the heritability of one population will not carry over to another. A few other comments on the estimation of heritability using the twin method are warranted, including comments on the assumption of non-additive genetic variance, the insensitivity of the twin method to gene-environment correlations and interactions, the low power of the twin method to detect heritability, and the assumption of equal environments.

N_d = the number of twin pairs doubled
N_s = the number of twin pairs not doubled, and
k = the number of regressors.

It was noted above that genetic variance may be additive or non-additive. However, the methods of calculating heritability described above, based on twice the difference between MZ and DZ correlations, assumes that all genetic variance is additive, because it assumes that DZ twins share 50% of their genes. In fact, although they share 50% of their additive genetic variance, they share only 25% of dominance variation (because both alleles at a locus must be the same for the same dominance relation obtain in both individuals), and probably even less of epistatic genetic variation, because all the alleles that interact must be the same in both twins for the same phenotype to emerge (this effect has been dubbed "emergenesis" by Lykken *et al.* 1992). To the extent that the genetic variance shared between DZ twins is less than 50% (and it always will be because some genetic variance is bound to be non-additive), heritability will be overestimated and shared environment underestimated. Although there are modeling techniques that attempt to identify non-additive genetic variance in twin data and separate it from additive genetic effects, the small number of subjects in this research currently precludes the use of these methods. Thus we will assume, as in most other twin studies of small to moderate sample sizes, that only additive genetic effects should be considered.

Another potential problem with estimating heritability using the twin method is the inability of the equations described in the last section to detect gene-environment correlations or interactions. The equations above are based on the assumption that these terms do not exist, so their effects are instead lumped in with heritability and environment. Although there are techniques that allow for the estimation of gene-environment interactions and correlations, they have low power (Wahlsten 1994) and require an experimental design with multiple environments to compare (Eaves 1982; Heath & Martin 1986), a design not used in the present study. Thus, we simply must bear in mind that the gene-environment interaction is another cause of unreliability in the estimate of heritability and environmentality provided by a twin study. These estimates should always be interpreted with caution.

Another unfortunate consequence of using the twin method to estimate heritability is that it has low power to detect heritability (Martin *et al.* 1978), thus requiring a large sample size and/or effect size for a significant heritability to be detected. Recall that heritability is based on the difference between two correlations. The power to detect differences between correlations depends both on the size of the difference and the values of the actual correlations, since differences between large correlations produce a larger effect size than the same difference between small correlations (Cohen 1988). To give the reader an idea of the number of pairs of twins needed to detect heritability at 80% power with a 1-tailed significance level of .05, the table below has a few examples using correlations of the magnitude that are typically found in twin studies of cognitive abilities. The studies described in the chapters that follow range in sample size from 8 pairs of twins (chapter 5) to 76 pairs (chapter 2). With these relatively small sample sizes, and with correlations that are relatively close in magnitude (as is the case in some of the studies to follow), the power of the DF method to detect heritability is not very large. Thus while I will provide h^2 and c^2 estimates from the DF method whenever possible, they may not register as significantly different from zero because of the small sample sizes.

Table 1.1

Sample sizes required for set heritability sizes and MZ and DZ correlations with power (ß) = 80% and a significance level of 95% (p = .05) 1-tailed.

h^2	r_{MZ}, r_{DZ}	total pairs of twins required
.2	.90, .80	80
.4	.90, .70	74
.2	.70, .60	350
.4	.70, .50	140

Finally, a last assumption of the twin method that also affects heritability estimates is the assumption that both MZ and DZ twins have "equal environments"—that is, that MZs do not have more similar environments than DZs. However, one might imagine that since MZ twins look more alike than DZ twins, and since they often have a shared placenta *in utero*, that they might have more similar shared environments than DZ twins. If MZs do in fact have more similar environments, but our model (in section 1.1 and 1.2) assumes that shared environmental variance is equal for MZ and DZ twins, then heritability would be overestimated. Since this is such an important assumption, the next section (1.4) is devoted to its consideration. To preview that section, though, there is some evidence to reassure us that MZ and DZ twins do have equally similar environments.

Because estimating heritability has so many pitfalls and because measuring it by finding the difference between two correlations has relatively low power, all of the estimates of heritability and environmentality provided in this work (and in other twin studies) should be interpreted with caution. Ideally, the same phenotypes would be studied in other individuals with varying genetic relationships (e.g., adoptive siblings, half-siblings) to see if the same outcome obtains. However, recall that the research presented in this thesis has the dual goal of not only estimating heritability but also *establishing* genetic effects on variation. Establishing that MZ twins are siginificantly more similar than DZ twins can be done with simple tests like t-tests with much higher power. Since a study of this detail has never before been undertaken with young children learning language, just establishing genetic effects is a novel contribution that justifies continuation of the present study and may inspire other studies in the same spirit.

1.4 Equal environments for MZ and DZ twins

The equal environments assumption mentioned in the previous section deserves a somewhat more detailed consideration, since it has been the topic of considerable research. This assumption must be tested at two points: in pre-natal and post-natal life.

In utero, two-thirds of MZ twins share one chorion (the outer layer of the placenta), while DZ twins never share a placenta (Plomin *et al.* 1990). Monochorionic twins usually share a circulatory system, which means they exchange nutrients, hormones, growth factors, and other compounds that affect neural and physical development. In addition, they are likely to share any pathogen that comes along. These factors can operate to make the environments of (two-thirds of) MZ twins more similar than DZ twins, violating the crucial assumption of equal environments. Indeed, Phelps, Davis & Schartz (1997) reviewed several studies of cognitive abilities, personality, and mental disorders that demonstrated greater similarity between monochorionic MZ twins than between dichorionic MZ twins. The inflation in MZ twin similarity, in the absence of any similar affect on DZ twins, could inflate heritability estimates.

However, the issue is more complicated, because just as sharing a placenta can make twins more similar, it can also serve to make them different. A shared circulatory system can result in asymmetries in blood flow, causing one twin to be deprived of blood and nutrients at the expense of the other. Both the deprived and the gorged twin can be damaged in the process if the sharing is sufficiently unequal. Typically, though, the smaller twin is more damaged. Scarr (1969) reported that within MZ twins, the smaller birthweight twins had lower IQ scores on average than their co-twins at ages 6-10 (though see Wilson 1985, where it is shown that the effects of birthweight on IQ dissipate over time until adolescence, where birthweight has little or no predictive value). Thus while sharing a placenta may increase MZ twin similarity, it can also server to make them different. At present, we simply do not know which force is stronger.

Ideally, all twin studies would separate MZ twins into mono- and dichorionic groups to see if their correlations differ. However, in a study as small as the present one, such separation is not feasible. Furthermore, it is not always possible to ascertain the placenta type of a pair of twins: sometimes placenta testing is not completed at the birth hospital or it is inconclusive. As more subjects are added to the present study, the topic can be investigated more thoroughly, and hopefully some of the larger twin studies already taking place elsewhere can investigate the issue as well.

The second arena in which environments for MZ twins might be more similar than those for DZ twins is in post-natal life. Parents and acquaintances of MZ twins may treat them more alike (consciously or not) because of their nearly identical appearance and in doing so may create a more similar environment for them (a type of gene-environment correlation). Research on this possibility indicates that although parents may in fact treat MZ twins more identically than DZ twins, this behavior may not have a noticeable effect on phenotypes under study. Loehlin & Nichols (1976) surveyed twins and their families on items including whether they dressed alike, played together (age 6-12 years), spent time together (12-18 years), and slept in the same room. They found that MZ twins did in fact have higher ratings on these items than DZ twins. To find out whether increased similarity on these items actually produced increased similarity between twins, Loehlin & Nichols restricted their analysis to MZ twins and correlated similarity on their survey items with similarity on 4 tests (National Merit Scholarship Qualifying Test, California Personality Inventory, Vocational Preference Inventory, and an interpersonal relationship inventory). They found no correlation between similarity ratings on their survey and similarity of performance on these tests (r ranged from -.06 to .06 (ns) with N ranging from 276 to 451 pairs for each correlation).

Scarr (1968; Scarr & Carter-Saltzman 1979) reached the same conclusion with a different method. In the process of doing a twin study of academic achievement in adolescents, it was discovered that about 200 pairs of twins in this study misclassified their own zygosity—that is, they reported that they were one zygosity (MZ or DZ) but a blood test revealed otherwise. Scarr and her colleagues were therefore able to compare the difference in correlations between twins of "true" and "perceived" zygosity. The results showed that similarities between twins were much better predicted by true zygosity than by perceived zygosity, meaning that DZ twins who thought they were MZ were no more similar to each other than other DZ twins, and MZ twins who thought they were DZ were as similar to each other as other MZ twins.

Thus, even if the treatment of twins is biased by appearance or perceived zygosity, it does not seem to make a difference in terms of similarity of phenotypes. The equal environments assumption is therefore perhaps the only one of the problems for the twin method considered in the last section that should not cause undue concern. Furthermore, where the topic of the present research—language—is concerned, it is probably reasonable to assume that learning environments are not significantly different for different types of twins. For such a difference to obtain, parents would have to systematically modify input to one twin as opposed to the other, an unlikely scenario in most cases, since twins typically spend every waking (and sleeping) moment in close proximity until they are old enough to walk. Any speech available to one infant twin should be available to the other, so that only a concerted effort by parents to separate their twins would produce different linguistic environments.

5.2 The generalizability of twin language to other children

Another important consideration in using twins is that they be representative of the population to which the researcher wishes to generalizes the heritability results. Indeed, it has been argued that twin language is qualitatively different from non-twin language. For a long time there was a pervasive assumption in the twin and language development literature that twins were somewhat delayed in language development and more prone to language disabilities, and even suggestions that they had secret languages (Zazzo 1960; Luria & Yudovich 1959; Day 1932; Davis 1937; Lytton 1980; Conway, Lytton, & Pysh 1980). As we shall see, though, any of these problems can also affect singletons; twins are just more prone to them due to their biological and social situation. Twins are thus not a completely different population whose results do not generalize to singletons.

The consensus in the literature of the inferiority of twins in language originates with two papers from the 1930's which remain two of the largest studies to date of the language of twins as a group. These are Day (1932) and Davis (1937). Day studied 80 pairs of twins and 140 singletons age 1.5 to 5.5. On several gross measures of language complexity (e.g., sentence length, number of different grammatical categories in a sentence) twins' spontaneous utterances were found to be as much as two years behind those of singletons. Davis brought the same method to twins age 5-9 years and found that on structural measures the twins caught up with singletons on average, but were still more likely to have articulation problems. However, there are several flaws in these studies (all previously noted by others). First, no effort was made to exclude twins who had language, speech, or hearing pathology. Since these are more common in twins (for reasons this will become clear below), it is possible that these deviant subjects lowered the average results for twins. Second, no information was reported on the birth weights or gestational ages of the twins, both of which may be factors in language delay. Third, the twins in these studies were not always observed separately, and the fact that two children of the same age were competing for the attention of the experimenter may have led to shorter sentences, making productive language appear delayed.

Subsequent studies have confirmed that on average, twins do lag behind singletons, but not by two years as Day claimed. Most studies find the language delay in twins to be 6-8 months or less (Mittler 1970; Hay *et al.* 1987; Koch 1966; McEvoy & Dodd 1992; Akerman & Thomassen 1991; Conway, Lytton & Pysh 1980). Furthermore, the studies available on older twins report that language delays subside by later childhood. Davis (1937) found 9-year-old twins to be nearly normal on her measures of syntax. Wilson (1975) found that 45% of the twins in his sample had Verbal IQs at least as high as their singleton counterparts. Furthermore, Record, McKeown & Edwards (1970) report that 18-year-old twins performed only about 5 points lower than singletons on a national college entrance exam (in the UK). And since all of these studies used an unselected twin population, it is possible that their findings reflect a few residual problems in the whole twin population, while most twins are within the average range.

While the consistent finding of some language delay in twins is troubling, it is important to consider the reasons for such delays and decide whether they are specific to twins or merely represent the same problems that all children potentially face. Several studies (e.g., Akerman & Thomassen 1991; Mittler 1970) suggest that biological risk factors such as low birthweight, decreased gestational age, and perinatal insult due to the delivery of multiples are responsible for cognitive impairments in young twins. On the other hand, Tomasello, Mannle, & Kruger (1986) showed that twin-singleton language differences correlated with a decrease in adult speech being directed towards each twin and fewer joint attentional situations between mother and twin compared to singletons, suggesting that twin language delay was also due to environmental factors. Savic (1980) and Malmstrom & Silva (1986) also showed that twins have fewer opportunities to talk than singletons and use shorter utterances, and they suggest that the environmental factor of having a constant interlocuter (i.e., the co-twin) results in these effects.

Three additional studies suggest twins suffer from both biological and social disadvantages. Lytton and colleagues (Lytton 1980; Conway, Lytton, & Pysh 1980) report that environmental variables such as maternal speech to child accounted for at least as much of the variance in language measures as biological variables (including birthweight, Apgar score , and time of gestation). Record, McKeown & Edwards (1970) examined children who lost their twin at birth or shortly after. In principle—that is, assuming that the death of the twin did not cause extra perinatal stress to the surviving twin—this group permits the separation of biological from social influences. They are biologically disadvantaged by being a twin (they shared resources in utero and were likely to have lower birthweight and gestational age), but they did not have to share social resources with a same-aged child. Record *et al.* found that these individuals performed intermediately: higher than the twin average but lower than the singleton average, suggesting biological factors cause some but not all twin delay. Finally, Mittler's (1970) results suggest that biological and social factors interact to disadvantage twins. Mittler found that birthweight, gestational age, and SES were significant predictors of twin-singleton language differences. However, twin-singleton differences only obtained for the lower and middle social classes, not the upper classes, suggesting that the same social disadvantages that all children are vulnerable to may affect twins more harshly. The interaction is probably due to the fact that lower SES families have diminished access to good pre- and immediate postnatal care, and the disadvantage is amplified in twins since they are already at risk biologically.

This summary indicates that there is language delay in twins on average, but that it is not extreme and it dissipates in late childhood. I submit that the use of twins is justified for several reasons. First, the delay in twins is not extreme but a matter of 6-8 months. Although means for twins at young ages will be slighter lower on language measures, it is not clear that their variance is different from normals. If the variance is similar in twins and non-twins, quantities like heritability, which are based on variance, should hold for both populations. Second, the early language delay clears up by late childhood, indicating that twins do not develop permanently deviant language but are merely slow starters. Third, one important cause of language delay, biological risk factors, can be eliminated by excluding twins who have low birthweights and gestational ages from study. Twins who develop language disorders later on can also be excluded post-hoc, which would presumably reduce the noticeable twin delay. Fourth, we can compare the performance of twins in this language study to established norms to uncover any delay that may exist and transform the data if possible or interpret it appropriately (i.e., as not generalizable to other populations). Taking these arguments and precautions into account, it seems reasonable to use twins to study the heritability of language all children.

1.6 The present study

The present study is the first step in applying the twin method to the development of language in order to establish genetic influences on variation and to estimate its heritability. If language development does have genetic influences, what can be concluded? While a finding of genetic influences on individual differences in language development would implicate a genetic role, it is not valid to translate the magnitude of heritability into the importance of genes in the development of language in each person. If language acquisition is, say, 20% heritable, this does not mean that genes are "20% important" in any individual. But we can conclude that genes must play some role in the development of language in any individual, in order for them to drive differences among people. Thus, we can learn two important things from studying genetic influences on variation in language development: (1) whether genes play a role in the development of language in any of us, and (2) the extent to which individual differences in language development are driven by genes. In fact, given the problems with the interpretation of heritability in the twin method reviewed in the last section, the first goal—establishing genetic effects—is probably the more useful and revealing one, as Wahlsten (1994) has also argued.

Another possible application of heritability in language development is in comparing aspects of language among themselves—e.g., making statements like "word learning is more heritable than grammatical learning." If we can trust our estimates of heritability, it would be interesting to eventually compare the heritability of language to other cognitive traits that have been studied extensively, such as memory and spatial imagery, and to non-cognitive domains as well, such as alcoholism and extroversion (Plomin, DeFries & McClearn, 1990). Making such comparisons could reveal whether language is dissociable from more general cognitive abilities, in terms of the causes of its individual differences.

Although this study represents the first attempt to study the heritability of the milestones of language development, one important caveat to keep in mind is that most of the phenotypes studied in this work are not specific to language. That is, although language is studied, very little effort has been made to partial out non-linguistic variables that may drive the results of this study. The only attempt in that direction has been to compare co-twins, when possible, on major milestones of motor development (such as rolling over, sitting up, crawling, and walking) and see how well those correlations predict correlations in productive language development. Thus, aside from partialling out the effects of very basic motor development, we cannot know if the genetic effects come from processes specific to language, or if they are driven by the heritability of more basic cognitive processes, such as auditory processing, conceptual development, or even overall brain maturation. The question is a philosophico-scientific one: can we imagine a more basic cognitive skill than language that might be responsible for its heritability (e.g., attention, short term memory, temperament)? Since previous scientific study of language has indicated that language cannot be reduced to these allegedly simpler components, I will for the time being assume that a finding of heritability for language as a whole is meaningful. However, as this study continues, more effort will be made to separate the heritability of non-linguistic precursors to language from that of purely linguistic abilities.

1.6. Preview

The next four chapters, 2 though 5, each present a twin study of the development some aspect of vocabulary or syntax. In Chapter 2, I present and defend the rather unconventional method to collect data used in chapters 2 through 4. I then review the twin literature on the development of vocabulary and describe the results of a vocabulary study with twins. Chapter 3 is devoted to a twin study of first word combinations. Chapter 4 includes a review of twin studies of grammatical development and presents the results of our past tense overregularization study. Chapter 5 describes a twin study on the development of verb tense and agreement over time, which uses a very different method from that of the previous three studies. Finally, in Chapter 6, I summarize the findings, review their implications and problems, and suggest directions for future research.

Chapter 2

Vocabulary development

2.1 Individual differences in vocabulary development and a review of twin studies

One of the most stunning examples of the uniformity of language acquisition, of the unfolding of a maturationally driven human instinct, is the child's passage through the earliest milestones of language. Children typically begin cooing around 3 months of age, babbling in language-appropriate syllables at 6 months, using their first genuine words around 12 months, and producing their first word combinations at about 18 months. Even children who are born deaf go through the same procession of early milestones, although their language deteriorates soon after if they are not exposed to a sign language (Goldin-Meadow & Mylander 1984). Observations such as these which show the uniformity and robustness of language in our species have helped to make the case for the innateness of language. However, despite this overall uniformity, there is some variance in age at each of these milestones. Since early word production is the focus of this chapter, the evidence for individual differences at that stage will be concentrated on here.

Variation in rate of learning has been observed as early the first 100 words Eighty percent of children produce their first words within 8-10 months of age, and while this is a small window, it should be clear that there is some variation. More variability appears as age increases. Bates, Dale & Thal (1995) report that at 16 months of age, children range from having produced no words (at -1.28 standard deviations below the mean) to 150 words (at +1.28 standard deviations above the mean), with a mean of 60 words. Furthermore, they found that the correlation between age and vocabulary between 8 and 16 months is .47, meaning that age accounts for $(.47)^2$ or 22% of the variance. Although this correlation is considerable, there are apparently factors other than just age that also account for differences in vocabulary growth.

The growth of vocabulary also has variation. The so-called vocabulary spurt, a relatively sudden increase in the rate of word learning typically seen at about the 50th word or around 18 months of age (e.g., Bloom 1973; Nelson 1973), may be present in some children and not others.

Goldfield & Reznick (1990), in a study of 18 children's longitudinal vocabulary development, found some children who displayed a spurt in word learning and others who did not. Furthermore, they reported a significant correlation between spurting and birth order: first borns were more likely to spurt than later borns. The authors attribute the difference to parents' having more time to play the "naming game" with first borns than with later borns, though one can imagine a host of other environmental differences between early and later-borns which might be relevant (more linguistic interaction with other children for later borns, more attention or input from parents for first borns, etc.) Individual variation in going through a vocabulary spurt was thus regarded as an environmental, rather than a genetic difference between children by Goldfield & Reznick. However, since the design of their study confounded genetic and environmental variation by studying only singleton children raised by their biological parents, the authors could not test this hypothesis directly.

In addition to individual differences in rate, differences in so-called "style" of word learning were first described by Nelson (1973) and have remained unresolved. Children whose first words are mainly comprised of concrete nouns have been dubbed "referential" or "analytic," while children who have a more even distribution of categories and who tend to imitate whole utterances unanalyzed before producing single words of their own have been dubbed "expressive" or "holistic" learners. These differences have been attributed to personality differences, differences in socio-economic status (SES) of the children's families, differences in maternal style, etc. (see Bates *et al.* 1995 for a summary). Some of these explanations are based on input or environmental factors (e.g., parental style, quality and quantity of language input, SES), while others are based on factors intrinsic to the child (e.g., personality differences).

What are we to make of these individual differences? As discussed in chapter 1, these differences can be used to learn more about the language development process. By using the twin method, it is possible to find out to what extent individual differences in language development are based in genetic differences and to what extent they are due to environmental differences. A finding of any genetic differences, or heritability, could support the argument for the innateness of language, since if variation in a trait is heritable, that trait must be coded for by genes. (Of course, as noted in Chapter 1, this study does not allow us to discriminate between the heritability of language-specific processes versus other cognitive abilities that may contribute to language.) Substantial environmental differences would also be informative about the process of language development. Such a finding would suggest that the rate and style of early word learning may be influenced by environmental factors that parents can take advantage of to speed their children along. Either way, armed with such knowledge, we might be better able to evaluate claims about individual differences in parenting such as those made by Goldfield & Reznick.

To date, no twin studies have addressed these issues specifically. Language-related behavior genetic studies tend to focus on verbal IQ, reading ability, or in a few cases, on other standardized tests of language. IQ is of questionable relevance to early language milestones. IQ is based on variation in developed intelligence, while the variation in learning words is only in the timing; the actual skill is the same in everyone (every normal person learns at least 100 words and can combine them into sentences of length greater than two). Studies based on standardized tests of vocabulary may not be relevant either, especially since they typically focus on a population that is older than the one we are interested in and, with one notable exception that will be considered in detail, they tend to measure subjects at one point in time rather than following them longitudinally. However, since such studies are the only existing evidence until now of the heritability of language, they are worth reviewing.

Overall IQ has been shown to be heritable in numerous twin studies of children and adults. Dozens of these studies were reviewed and analyzed thoroughly in Bouchard and McGue, 1981 (along with studies using other behavior genetic methods). They report an average MZ correlation for overall IQ of .86 for MZ twins and .60 for DZ twins, yielding a heritability of about .52. Verbal IQ alone is also heritable. In a review of four studies of verbal IQ in subjects ranging in age from about age 4 to 6 (Segal 1985; Wilson 1975; Koch 1966; Stevenson *et al.* 1987), Stromswold (under review) finds the average MZ correlation for verbal IQ is .82 while the DZ correlation is .49, yielding a heritability of .66.

Nichols (1978) reviewed studies of more specific verbal abilities in adults and children, including verbal comprehension and verbal fluency subtests of a verbal IQ battery. For verbal comprehension, collapsing across 27 studies, he reports an average MZ correlation of .78 and an average DZ correlation of .59, for an estimated heritability of .38. For verbal fluency, averaged over 12 studies, Nichols reports an average MZ correlation of .67 and an average DZ correlation of .52, yielding a heritability of .30. While these reviews yield somewhat different estimates of the heritability of verbal abilities (.66 for overall verbal IQ versus .38 and .30 for specific components), they both implicate a large heritable component in these abilities.

Cross-sectional twin studies specifically of vocabulary skills have been reported by Segal (1985), Foch & Plomin (1980), Thompson, Detterman & Plomin (1991), Mather & Black (1984), and Fischer (1973), all of whom used standardized vocabulary tests such as the Stanford-Binet vocabulary subtest (Terman & Merrill 1960), the Peabody Picture Vocabulary Test (PPVT) (Dunn 1965), the Mehrabian Vocabulary test (Mehrabian 1970), and the WISC vocabulary subtest (Wechsler 1974). The twins in these studies ranged in age from 2.5 to 13 years and the sample sizes ranged from 21 pairs of twins (Fischer) to 146 pairs (Thompson *et al.* 1991). All these studies report significant or near-significant differences between MZ and DZ correlations. A meta-analysis of these and other studies estimating MZ and DZ correlations for "spoken vocabulary" is reported by Stromswold (under review), who finds the mean correlation coefficient is .73 for MZs and .44 for DZs (h^2 = .58). Thus there appears to be a heritable component to performance on vocabulary tests between ages 2 and 13.

These studies document the heritability of mature language skills in adults and children, but for the most part they fail to access the language development process in its earliest stages, the first 3 years. Furthermore, although verbal IQ has some relevance to language, it is also heavily dependent on general cognitive abilities as opposed to language per se. There is one series of studies that does address very early language development. It is part of a larger study of early intelligence based on the MacArthur Longitudinal Twins Study (MALTS) sample and has been published in reports by Plomin *et al.* (1993), Reznick (1996), and most recently by Reznick, Corley & Robinson (1997). These researchers sampled a host of emotional, temperamental, and cognitive measures for about 200 pairs of twins (100 MZ and 100 DZ) at 14, 20 and 24 months. Their language measures included the language-related items from the Bayley Scales of Infant Development (BSID; Bayley 1969), the Sequenced Inventory of Communicative Development (SICD; Hedrick , Prather, & Tobin 1975), and a vocabulary test using a looking time technique to assess comprehension (Reznick 1990). The SICD is designed to measure communication broadly. Its expressive scale includes items like "Name picture of a baby" and "Answer question *yes*" and its receptive scale has includes the items "Respond to *don't touch*" and "Indicate referent *ears*." The BSID also appears to be a crude measure of communicative ability. Its expressive component includes the items "Vocalize 4 syllables," "Use words to make wants known," and "Name 1 object." Its receptive component includes "Respond to *No no*" and "Indicate referents for body parts on doll" for younger subjects, and for more advanced children it has items such as "Respond to 2 prepositions" and "Respond to *put 1 block on the paper*."

Reznick *et al.* (1997) seems to be the most recent report of the language results of the MALTS study, so the results reported there will be the focus of this summary. The BSID and SICD verbal components were combined to make two language constructs, one measuring expressive language and the other receptive language. Although the linguistic component of these measures is difficult to tease apart from what surely must be substantial effects of shyness and attention to experimenters, the constructs are worth examining, since they are the first attempt to measure the heritability of language in children so young. The expressive construct had a heritability (h^2) of .01 at 14 months (with shared environment, c^2, of .35); h^2 of.25 at 20 months ($c^2 = .49$); and, h^2 of .38 at 24 months ($c^2 = .40$). The receptive construct showed $h^2 = .28$ at 14 months (with $c^2 = .34$); $h^2 = .13$ at 20 months ($c^2 = .52$), and $h^2 = .18$ at 24 months ($c^2 = .51$). Reznick's word comprehension measure, based on looking time (a laboratory procedure), yielded extremely low heritability and shared environment: at 14 months $h^2 = .00$, $c^2 = .07$; at 20 months $h^2 = .00$, $c^2 = .22$; and at 24 months $h^2 = .08$, $c^2 = .17$. However, this measure had extremely low correlations between both types of twins at 14 months (.06 for MZ and .07 DZ) and rather modest correlations at 20 and 24 months (around .25 for all). These results would seem to suggest that non-shared environment ($1 - h^2 - c^2$) plays an important role in word comprehension. This conclusion is puzzling, since the only words children can know are those used in their environment, and twins share their environments to a great extent. It is more likely that something in the procedure itself is driving this result, since error tends to get put into the non-shared environment component.

Although there is not much to be said about the puzzling word comprehension results, three important points can be made about the other data. First, although it is small, there is heritability for both these constructs (expressive and receptive language) at almost all points of measurement. (The first point of measurement, 14 months, was not heritable for productive language, perhaps because production has barely gotten off the ground at this age). This finding is encouraging, because even though the broad measures of communication used in Reznick *et al.*'s study are somewhat different from the more specifically linguistic skills in the present study, they represent pre-requisite or concomitant skills, so it would not be surprising if the heritability of word production were similar. Second, there is significant shared environment at all points of measurement—much larger than heritability. This result is not surprising, since the specific words and phrases that children recognize and use can only be those they have heard, and these are bound to vary somewhat by family. The same effect should be present in any study of language unless a pure measure of linguistic ability, independent of actual language, could be used. Third, for language production, heritability increased over the three measurements. The increase could merely be due to the relative absence of productive language at 14 months compared to 20 and 24 months. Supporting this idea, the variance of the production scale increased after the first measurement (s.d. = .12 at 14 months, .23 at 20 months, and .20 at 24 months) while for the receptive scale the variance remained steady (s.d. = .14, .18, and .14 at the same points). On the other hand, the increase could actually reflect increased gene effect or new genes coming on line. A Cholesky decomposition (or triangular factorization) performed by Reznick *et al.* indicated that new genetic effects emerged for expressive language at 20 months and the same genetic effects were responsible for heritability at 24 months.[6] This model implies that genes that affect expressive language are present but lie dormant until 20 months. As we shall see later in this chapter, a hint of this increase in heritability over time comes through in the present study, which is also based on production.

The cross-sectional studies of vocabulary (Foch & Plomin 1980; Thompson *et al.* 1991; Mather & Black 1984; Fischer 1973) were useful inasmuch as they indicate the heritability of vocabulary at later stages of development. Although these results are interesting in themselves, they cannot go very far in answering questions about the rate and structure of very early word learning because the subjects used are far beyond this early stage and because testing is not longitudinal. The longitudinal study by Reznick, Corley & Robinson (1997), is a step in the right direction and certainly the most comprehensive and ambitious study of the heritability of the development of cognitive variables. However, it still cannot answer much more specific questions about rate and style in vocabulary acquisition like those discussed above because it has only three points of measurement and because the language measures are rather cursory. Thus, despite the existing body of work on the heritability of vocabulary development and skill, there is still a need for more research to answer the kinds of questions we raised in the last section.

The goal of the present study is to make a contribution to the body of literature on the heritability of vocabulary by providing the kind of data that have not previously been gathered for twins: long term, high-sampling rate data from early word production. Armed with these data, it may be possible to examine questions about the source of individual differences in rate and style described in the last section.

2.2 Method

Because the method used in this study to follow vocabulary development is not typically used in large studies with untrained parents, this section is devoted to a rather lengthy description and defense of the methodology. Many of the issues covered here are relevant not only for the

[6] The Cholesky decomposition for heritability is based on the comparison of genetic variances at the various points

vocabulary study, but also for the word combination study (Chapter 3) and the overregularization study (Chapter 4), since the data for the latter two studies were also gathered by parent report. However, I will focus on the vocabulary study since more data are available on the reliability of the methodology for vocabulary than for the other measures.

The development of productive vocabulary was tracked by asking parents to keep diaries of their twins' word use. Parents kept a list of all the different words used by each twin each day, even if the words had been said on previous days.[7] These daily lists were kept on forms which the lab provided. The forms had spaces to write in the date, each twin's name, each twin's daily word list, comments on the child's health, and a few other details. (See Appendix B for copies of the forms.) Participants mailed these journals back to the MIT Twins Study laboratory in Cambridge, Massachusetts each week in stamped envelopes provided by the lab. They were contacted by phone at regular intervals (3-4 weeks) to address problems and questions that arose during the course of the study.

Parents began keeping vocabulary records at their twins' first words whenever possible (i.e., when they joined the study early enough) and continued each day until their twins reached about 100 words, or for as long as they were willing. They were told that an utterance was to be counted as a word if it was used consistently to refer to one thing or class of things and not to others. They were also told to note which words were imitations and which were spontaneously produced, and to write down any accompanying context if it might help elucidate the meaning of a word. In cases of uncertainty over whether to count an utterance as a word, parents were told to use their intuition and to be as consistent as they could from day to day and from one twin to the other. (For the exact written instructions given to parents, and an example of a completed form, see Appendix B.)

Despite the careful instructions, there may be some concerns with using parent report. After all, can parents be trusted to keep accurate records? Although parent-based measures are

---

in time to see if they are likely to be produced by the same genetic factors.

fairly common in vocabulary studies (e.g., Fenson *et al.* 1994; Goldfield & Reznick 1990; Lieven *et al.* 1992) , the one used here is uncommon on a large scale. In the remainder of this section three classes of problems with this method will be addressed. The first is the use of parent report in general. The second is the use of a free-report journal instead of a recognition-based language inventory. The third is the use of parent-report with twins in particular.

In trying to defend the method used in this study, the existing critical literature on parent report in language development will be reviewed and three additional procedures, which were undertaken to assess parents' accuracy, will be reported. (1) Subjects completed a MacArthur Communicative Development Inventory (CDI: Words and Sentences) for each twin at least once during the study and these were compared to the parent's journal records. (2) Four pairs of MZ twins were videotaped for an hour a day for 4 to 5 days and a research assistant generated a journal page based on each of the tapes to compare with the parent's reports for those days. (3) For 10 pairs of twins, more than one person happened to keep the daily journal. The comparison of two different journal-keepers for the same child, which was unintentional but fortunate, can reveal how idiosyncratic a particular observer is and is especially to assess whether rater bias, which inflates shared environment, is a problem.

2.2.1 Parent report in general

Parent report was necessary in these studies for two reasons. First, due to constraints on time and staff, it would have been impossible to arrange for laboratory or home visits often enough to track children's language development on a daily basis. Second, due to power considerations covered in Chapter 1, it was necessary to amass as many twins as possible and being limited to the Boston, MA area would have made this goal difficult. Twins comprise about 1/80 of live births among Caucasians, and since 1/3 of those are different-sex fraternal twins, only 1/120 were useful for the study. Restrictions on birthweight, gestational age, and other health factors (see Subjects

[7] In some cases exceptions were made. For instance, if a child had used the same words for many days in a row (e.g., *mama* and *dada*), the parent was told that s/he no longer needed to write them down. Exceptions like these were made to prevent parents from getting too bored or irritated with list-keeping.

section below for more details) further decreases that number. Furthermore, since this study was labor-intensive, many parents of twins do not even respond to ads for the study and even some of those who do respond ultimately choose not to join the study. So, a parent report method allowed for the recruitment of twins from all over the United States and Canada, which greatly increased potential sample size.

Aside from allowing a larger subject pool, parent report has a long tradition in studies of early vocabulary production because it has some benefits. It is probably more accurate than occasional experimenter observation, since parents have a larger range of contexts in which to observe their children's language. Furthermore, a child's talkativeness is highly influenced by his temperament, so a child who is fearful of strangers is less likely to give them an accurate sample of his language ability. The developers of the MacArthur CDI, an instrument developed for parent report of language development, have shown it to have high test-retest reliability (r = .8 to .9), high internal consistency between different vocabulary measures within the inventory (r = .95 to .96), and high reliability when compared to laboratory measure of productive vocabulary (r = .72) (Fenson *et al.* 1994; Bates, Bretherton, & Snyder 1988), lending further support to the notion that parents are good recorders of their children's language development. For all of these reasons, parent report of one kind or another has been used extensively in the early language development literature.

Despite these advantages, there is some concern that parents may bias their estimates of their children's vocabularies in certain ways. Pine, Lieven & Rowland (1996) compared vocabulary estimates made by experimenter observation (in a single recording) and parent report (using a combination of a checklist inventory and a free-report list). They found that when they equated for total vocabulary, the proportion of common nouns was greater in the parents' reports than in experimenter observations by about 20%. The authors speculate that the truth lies somewhere in the middle: checklists lead to an overestimate while observations, due to their restricted context and temperament effects, lead to an underestimate. Their concern is relevant to this study, though. Since this study aims to examine of the proportion of nominals in a child's vocabulary, the difference between the two methods suggests that more than one method should be used to accurately estimate the proportion of different semantic categories in a child's vocabulary. However, we will show, using data from videotapes, that our method is not prone to a nominal bias as Pine *et al.*'s parent report inventory was. The difference probably comes from our measure being a free-report, daily diary rather than an occasional checklist inventory.

As their parent-report measure, Pine *et al.* used a checklist inventory combined with a small amount of free report which was collected at two points in the child's development (approximately 50 and 100 words total vocabulary). The parent's report is necessarily retrospective. For their observational measure, they used recordings taken at the same points. In contrast, the method used here combines the advantages of both these methods: it is recorded by the parent, who has access to a greater range of contexts and moods of the child, and it is not retrospective. If the common noun bias of parents comes from retrospection (names for things are easier to remember than other parts of speech because they can be cued visually for the most part), the method of the present study eliminates that problem.

Our evidence that our method does not share the noun bias of Pine *et al.*'s parent report data comes from a videotaping project. Parents' journals from four families were compared to videotapes one hour long made of their twins on the same day. The twins were videotaped at an average age of 22 months, with a range of 20 to 25 months. They had an average number of 104 words, with a range of 40 to 201 words, according to the cumulative journal records parents had sent in. The families were chosen because the twins were still saying only single words (few or no word combinations), they were identical (see section 2.2.3), and most importantly they lived in the greater Boston area (within a 60-minute drive with city traffic). These four families exhausted the number who met those criteria from the entire sample.

Each family had at least 5 videotaping sessions over the course of 2-3 weeks so that they would become accustomed to the author being there with the camera. The parents were told that the purpose of the videotaping was to learn more about how the twins interacted and used language with each other and to provide more context for their language. The parents were also told to continue to keep their daily lists for the period being videotaped because it would be hard for observers in the lab to tell what the children were saying without the parents to help decode it (the children were between 1 and 2 years old and their pronunciation was often quite idiosyncratic, as with most 1-year olds). Parents were debriefed at the conclusion of the final taping session. Despite these instructions, some of the families did not keep lists for all of the sessions taped. Two of the 4 families had both parent and video records for 5 of 5 taped sessions, but one had both video and parent records for only 4 of 5 taped sessions, and another had complete records for only 4 of 7 sessions.

The videotapes were watched by 1-3 research assistants (sometimes separately, sometimes together), who generated a word list for each session. These observers were permitted to watch the tape as many times as necessary and to use the parent's journal to help them decipher the children's words. The goal was for the experimenter's list to be as accurate as possible a representation of the words the child actually said, rather than just another real-time journal like the parents'.

Although videotaping twins and their parents may not be the best way to capture their natural behavior, it can provide some insight into how accurate parents' records are. It was our hope that repeated videotaping of the same families would make them more relaxed and natural as the sessions went on. Furthermore, since parents were not aware of the true purpose of the taping, they may not have made a special effort to keep accurate lists. In fact, in 2 of the 4 families, the parent (or caregiver) did not appear to be keeping word lists at all while I was there, apparently planning to write them down later. However, there is no way to know for sure how unnatural a situation the camera created, so the data should be viewed with some caution.

To compare the types of words that were noted by each method, the word lists from all the videotape sessions for each child were combined into one estimate of that child's vocabulary, and all the parent's journal entries for the corresponding days were also combined to make an estimate for the same child. Table 2.1 below shows the overall estimate for each twin from parent journals and experimenter's observations. (Families are named by their first initials.) As the table makes clear, the two estimates were quite similar. The overall correlation between parent estimates and experimenter estimates was .95. Table 2.2 shows the proportion of nominals for both raters. The correlation between parent and experimenter estimates of the number of nominals was .96. Furthermore, there does not appear to be a trend in terms of one method making higher estimates than the other. In two of the families ("F" and "O"), the parents' estimates were an average of 11% and 5% higher than experimenter estimates, respectively, but in the other 2 families ("N" and "C"), parents' estimates of nominals were 6% and 5% lower. Thus if these videotapes are to be trusted, they appear to show, that Pine *et al.*'s finding that parents overestimate nominals by an average of 20% is not true for the parent report method used in the present study.

Table 2.1

Parent and experimenter estimates of the number of words used by four pairs of twins during videotaped period

	Twin A: parent estimate	Twin A: video estimate	Twin B: parent estimate	Twin B: video estimate
F family	77	62	35	37
N family	63	67	39	38
C family	22	19	12	18
O family	37	41	41	46

Table 2.2

Parent and experimenter estimates of the proportion of nominals used by four pairs of twins during videotaped period

	Twin A: %nominals parent	Twin A: %nominals video	Twins B: %nominals parent	Twin B: %nominals video
F family	0.71	0.65	0.83	0.68
N family	0.29	0.30	0.28	0.37
C family	0.05	0.00	0.00	0.17
O family	0.43	0.37	0.46	0.43

2.2.2 Free report versus a checklist

As should already be apparent from the discussion in the last section, there are two types of child vocabulary reports used with parents: a periodic, free-report journal like the one used in this study, and an inventory which allows parents to choose the words their children know from a long list. The former is typically used by psychologists keeping track of their own children's development (e.g., Dromi 1987), while the latter has more commonly been used with untrained parents. One inventory developed for untrained parents is the MacArthur Communicative Development Inventory or CDI (Fenson *et al.* 1994). As we saw above, the developers of this checklist have shown it to be fairly accurate in estimating a child's vocabulary (compared to periodic laboratory testing) and reliable.

The obvious choice, then, would have been to use an inventory. This option was rejected for two reasons. First, frequent records were needed to follow the vocabulary spurt and the course of overregularization. The CDI (which has both a word and a grammar inventory) was not developed for frequent use, but for a one- or perhaps two-time assessment of the same child. It is not clear how reliable using the CDI as often as, say, once a week, would be. A related problem is that the inventory takes quite a bit of time some to fill out, since the parent must search the entire list to find the words the child has said. Parents may tire of filling out the CDI so often (and keep in mind a separate one has to be completed for each twin).

Second, the CDI was developed for parents with one child of the appropriate age, not twins. Parents of twins might confuse which twin had said which words in a retrospective inventory (this in fact turned out to be true; see below). A free-report daily journal solved both these problems. Being parent-based and high sample rate, it is privy to a wide range of contexts (and furthermore it is not as affected by temperament as experimenter observation). Since parents write down words each day (rather than having to remember back over weeks or months), they are less likely to confuse co-twins. Furthermore, as we saw in the last section, the journal is probably less biased towards particular categories, since if there is category-specific memory degradation, there is less time for it in a daily journal.

Fortunately, Reznick and Goldfield (1994) provide some evidence that diaries, even when kept by untrained parents, are nearly as accurate as checklists. With 24 (singleton) subjects measured at 2-month intervals from 1;2 to 1;10, they showed that the word estimates provided by parents' diaries were highly correlated with checklist estimates made at each age. Correlations ranged from .60 to .86. The only troubling result was that the correlations showed a slight decreasing trend over time, prompting Reznick & Goldfield to suggest that diaries were more accurate for lower levels of vocabulary than for higher levels. However, the number of subjects in this study also decreased as the number of words increased so that at the last time of measurement, only 7 children remained. Thus more data should be collected to confirm this suggestion.

To confirm that checklists and diaries would have provided similar estimates for this study, and to compare the proportion of nominals given for each, parents of 41 pairs of twins already participating in the present study filled out CDIs once in addition to keeping their journals. Seventeen of these pairs are also in the sample considered later in this chapter; the rest were still in the process of keeping journals when this thesis was being written and so have not been analyzed yet, but they were recruited and instructed in the same way as the other subjects so we can assume they are representative. These 41 parents were told to skip their journal for a day and fill out a CDI for each twin instead, following the printed directions. The CDIs were then entered into a database to be compared to each twin's accumulated vocabulary journal. The CDI was only used for the vocabulary study; it did not lend itself to comparison of the word combination or past tense journals.

The CDI captured some words that never showed up in parents' journal entries. This makes sense since in a recall situation, some words may escape the parent's memory, but an actual list may jog it. On average, about 37% of the words on the CDI for a particular child had not been reported in that child's journal. However, a journal has its advantages, too. An average of 37% of the words from the accumulated journals were not on the CDI. Presumably this is partly because the journal is not retrospective, so there is less time to forget words a child has said, and partly because not all the words a child says appear on the CDI; it is a fixed list. (Marcus *et al.* (1992) also noticed this effect.) Overall, 61% of the subjects were within ±10 words whether estimated by CDI or journals, and 80% were within ±20 words. Neither method was consistently higher: 37 of 82 children had higher CDI estimates, 42 had higher estimates by the journal, and 3 had no difference. The correlation between number of words as estimated by CDI and the number as estimated by journals was $r = .98$. So, at first glance, both methods appear to yield a similar estimate of vocabulary size. Presumably, true vocabulary size is closer to the union of the two estimate (though see the data on twin confusion in section 2.2.3.1).

However, the two types of report differed as to the estimates of content, with the checklist inventory yielding a higher estimate of nominals than the journal. (Nominals are defined as any noun except proper names, which are considered to be more social than nominal in origin by the inventors of this distinction (Nelson 1973)). The proportion of nominals in the CDI and in the journals was checked in one randomly selected twin of each of the 18 pairs who had between 25 and 235 words when the inventory was completed (mean 84 words by both CDI and journals). The proportion of nominals in the CDI report (not including proper names) was 58%, while the proportion of nominals in our journals up to the same date was 47%. All but two children followed this trend. Since the CDI obviously lacks many of the specific proper names children learn, I did two more analyses: one including proper names as nominals, and another taking proper names out of consideration altogether. In both of these analyses, the same trend held: the CDI had a higher proportion of nominals than the journal. In all three cases, the difference was significant by (paired) t-test ($t = 4.2$, $p = .0005$ for proportions of non-proper name nominals; $t = 1.74$, $p = .05$ with proper names included as nominals; $t = 2.79$, $p = .006$ with proper names excluded altogether). The lower proportion of nominals in our journals is encouraging, because like the videotape comparisons we made, it suggests that Pine *et al.*'s finding of a noun bias in parental report is specific to inventory report and is alleviated by our method.

Thus, in addition to meeting the practical necessities, the free report journal provides a similar estimate of vocabulary size to the CDI (confirming the findings of Reznick & Goldfield 1994) and also appears to be less vulnerable to the noun bias (allaying the admonitions of Pine *et al.* 1996).

2.2.3 Parent report issues specific to twins

Several more serious problems arise specifically with respect to twins. The first is twin confusion and the second is observer bias.

2.2.3.1 Twin confusion

Parents may be able to keep good records of the language development of one child, but may confuse which twin says which words. This possibility is troubling for two reasons. First, the worst case scenario is that MZ twins' words will be confused more often than DZs'. Assuming that MZ confusion is random (i.e., not systematically biased towards one twin), such confusion would make MZs appear more similar than they really are, artificially inflating heritability. Second, even if MZs are not confused more than DZs, equal confusion of both types would increase correlations for both and artificially increase estimates of shared environment. While this effect would not affect the heritability estimate (recall that heritability is twice the difference between MZ and DZ correlations), it would decrease the estimate of non-shared environment (e^2) and increase shared environment (c^2), since if h^2 is constant, e^2 and c^2 are a function only of r_{MZ}. Both effects were investigated in two ways: (1) by comparing the CDIs and the parents' journals for twin confusion and (2) by looking for twin confusion by parents (compared to experimenter observations) in the videotapes. As a result, fears of twin confused were attenuated.

We noted above that one reason for using free report instead of an inventory like the CDI was the CDI's potential to promote confusion between twins because it is a retrospective inventory. This fear turned out to be warranted, as a comparison of the journals and the coincident CDIs revealed. The CDI confusions were useful to us, though, because they give us the opportunity to examine the kinds of confusions that are made, and because they validate our decision not to use the CDI as our main instrument of collecting data.

CDI confusions were discovered because, of the words that showed up on the CDI but not on the accumulated journals for a particular child, an average of 20% appeared *in co-twins' journals*. In other words, if we can trust the journals, the parent checked off words on the CDI for one twin that had really been said only by the other twin, and this accounts for 1/5 of the words that the CDI appeared to capture over the journal. While it is possible that the retrospective CDI is more accurate, and there is no way to decide, it seems likely that the parent had more confusion between twins' words on the CDIs than in the daily reports.

While the finding of some twin confusion in the CDI is troubling because it raises the concern that parents may be confusing their twins in their journal reports as well, it did provide one encouraging finding: there was no MZ-DZ difference in the confusion effect. Parents of MZ twins did not confuse their twins' words on the CDI more often then parents of DZ twins. The MZ mean proportion of confusions was .21 (n = 46 pairs), the DZ mean was .20 (n = 18 pairs). This result is encouraging, since it implies that even if there is some twin confusion in the journals, it probably is not greater for MZs than for DZ. So the worst case scenario, in which heritability would be inflated by increased MZ correlations, appears to be unlikely.

While the journals seem to be less vulnerable to twin confusion than the CDI, there may still be twin confusion in them. To investigate this possibility, the videotapes that were also used in section 2.2.1 were reviewed again. Since resources were limited and the possibility of greater MZ than DZ twin confusion was the largest fear, only MZ twins were videotaped. The reasoning in taping only MZ twins was that if these tapes showed little or no confusion, the fear of inflated heritability due to MZ confusion could be put to rest.

With the videotapes, two possible types of twin confusion can be checked. The first type occurs when twin A says a word and twin B does not, and the parent fails to write down the word for twin A and instead writes it for twin B (or vice versa). Such mistakes can be called "reversal errors" since they represent a complete reversal of attribution. The second type of confusion occurs when twin A says a word and twin B does not, but the parent writes it down for both twins. We can call these mistakes "spreading errors" since attribution appears to spread from one twin to the other. In both cases, the videotapes cannot give us definitive answers, since they are only one hour long and a child may have said a word off-camera that the parent wrote down. However, if we assume that the camera captured all words and the parent only kept track of the words on camera, then comparing the video reports and the parents' reports can provide an estimate bound for the real number of confusions.

To find the number of the reversal errors, the parents' list for one twin was compared to the experimenters' lists for that twin, looking for words that appeared on the parent's list but not the experimenter's. These are the parents' "extra" words. I then compared the same lists for the co-twin, this time looking for words that appeared on the experimenter's list but not the parent's. These are the experimenter's "extra" words. Finally, I compared the parent's extra words for the first twin to the experimenter's extra words for the second twin, looking for any overlap. The process was then repeated for the other twin, and done in all four families for all the sessions, yielding a total of 36 test cases if each twin is counted separately.

The proportion of the reversal errors for each session for each twin is shown in the second and third columns of Table 2.3 below. This proportion represents the number of words the parent wrote down for the child that should have been attributed to the child's co-twin instead (according to the system described in the previous paragraph). As the table shows, there were a small number of reversal errors. The worst case was as high as 17%, but the overall average was just 2%. While it would be preferable to rule out reversals altogether, it seems they are at least uncommon. Of course, these data must be taken with two caveats, both of which were already reviewed above: parents may have increased their accuracy for the camera and the twins may have said some of these words off-camera. In the first case, the numbers here would be underestimates of the truth, in the second case, overestimates.

To find spreading errors—where one twin's word is recorded for both twins—the parent's list for one twin was compared to the experimenter's list for that twin to look for words that overlapped. The parent's lists and the experimenter's list for the other twin were then compared, looking for words that the parent recorded but the experimenter did not. The overlapping words (between parent and experimenter lists) from the first twin and the non-overlapping words from the parent's list of the second twin were then checked for overlap. The process was then repeated for the other twin, and done in all four families for all the sessions, again yielding a total of 36 test cases if each twin is counted separately. The proportion of spreading errors for each twin is listed in the last two columns of Table 2.3. This proportion represents the number of words the parents wrote down for that child and his twin that were really only said by the twin (according to the system described above). Although there were some disturbingly high proportions of this type of confusion (see the second twin in the "F" family), the overall average was just 7%.

Table 2.3

	Reversal errors		Spreading errors	
	Twin A	Twin B	Twin A	Twin B
F family				
Session 1	0.00	0.00	0.00	0.33
Session 2	0.00	0.07	0.06	0.47
Session 3	0.00	0.00	0.07	0.09
Session 4	0.00	0.00	0.09	0.11
Session 5	0.00	0.07	0.10	0.21
F average	**0.00**	**0.03**	**0.06**	**0.24**
N family				
Session 1	0.17	0.11	0.00	0.11
Session 2	0.00	0.00	0.00	0.22
Session 3	0.06	0.11	0.06	0.11
Session 4	0.04	0.00	0.00	0.00
Session 5	0.00	0.00	0.05	0.05
N average	**0.05**	**0.04**	**0.02**	**0.10**
C family				
Session 1	0.00	0.00	0.00	0.14
Session 2	0.00	0.00	0.09	0.20
Session 3	0.00	0.00	0.13	0.00
Session 4	0.00	0.00	0.00	0.00
C average	**0.00**	**0.00**	**0.05**	**0.09**
O family				
Session 1	0.00	0.00	0.00	0.07

Session 2	0.00	0.00	0.00	0.13
Session 3	0.00	0.00	0.15	0.07
Session 4	0.08	0.00	0.00	0.06
O average	**0.02**	**0.00**	**0.04**	**0.08**
Overall average	0.02		0.07	

Unfortunately, since no DZs were taped, we cannot say for sure yet that MZs were not confused any more than DZs, though the CDI confusion data are encouraging in that respect. Furthermore, we have not been able to rule out twin confusion generally: clearly twins are being confused, if only in small amounts. Even if MZ and DZ confusions are equal, they will drive up estimates of shared environment. Until more videotapes of both MZ and DZ twins are made, we cannot be sure how large this effect will be. Therefore, when estimates of shared environment are made in this study, they are perhaps better viewed as upper limits than accurate approximations.

2.2.3.2 Observer bias

Finally, another concern for this method is observer bias. This factor drives up shared environment. Observers necessarily introduce some bias that makes their observations more similar and since a pair of twins shares the same rater, their correlation may be inflated not by their true similarity but by their observer's bias. This effect increases both MZ and DZ correlations and hence inflates shared environment estimates (although in principle is does not affect heritability unless bias is greater for MZ than DZ twins). Rater bias has been found in parent-report twin studies of personality and temperament (Neale & Stevenson 1989; see Neale and Maes, in prep, chapter 11, for discussion). It was also found with the developmental inventory used by Reznick, Corley & Robinson (1997). Reznick and his colleagues found that, compared to experimenter administered versions, parent-administered versions of both the SICD receptive and expressive inventories showed significantly higher co-twin correlations at 14 months (regardless of zygosity), at 20 months (with more inflation for DZs than for MZs), and at 24 months (with ambiguous zygosity effects). Parent-generated correlations were higher in every case, but the margin varied from as little as .03 to as much as .49. For expressive items, which concern us the most since this study is based only on observed production, parent inflation of twin similarity was highest at 14 months and decreased only moderately by 20 months. Since the twins in the present study were observed at this same age range, these results are cause for concern.

There are two reasons not to panic about the bias reported in Reznick *et al.*, though. The first is that the SICD is a retrospective inventory. It may be more prone to bias than a daily journal because events have to be not only interpreted but also reconstructed, adding an extra opportunity for bias. The second is that the items on the expressive SICD are likely to be influenced by temperament. They are:

22a. Name picture of baby

22b. Name picture of shoe

22c. Name picture of ball.

29a. Answer question "No".

29b. Answer question "Yes"

31a. Answer question "shoes"

While an actual test of these questions would, in principle, provide a more accurate assessment than a parent's guess of whether they are true, assessment by a stranger would be highly susceptible to temperamental factors, like refusal to participate or distractedness by a new person. This effect might make the twin correlations provided by the "objective" observer lower than their true value, meaning that parents are not inflating twin similarities as much as Reznick *et al.* suspected. The true twin similarity probably lies between the objective observer's and the parent's estimate.

In the present study, the unexpected opportunity was provided to compare estimates made by different observers.[8] For some twins, more than one person kept the journals, and vocabulary estimates from two different observers of the same pair of twins could be compared. Typically,

---

[8] Special thanks to Sonia Chawla, who single-handedly prepared and analyzed these data.

one observer was the mother and the other was a grandmother, the father, or a daycare provider. In order to qualify for this analysis, the two raters had to have kept the journals on different days (to minimize collaboration), they had to have a minimum of 10 days of journal entries each, and there had to be significant interleaving in their journal-keeping days, so that we could reasonably expect the estimated number of words to be the same. Of course, since the journals were recorded by the different observers on different days, there will be some differences in the actual words. If the different observers (for example, a mother and a nanny) participate in different activities with the twins, or if the twins are more talkative with one person than the other, then the actual words used by the twins will be different and comparing the two observers is not fair. However, given enough time and sufficient interleaving of days, the overall number of words should be similar.

To ensure that the two observers were keeping records around the same period of time, the dates of their journal entries were separated and equal numbers of days were taken from both raters. These days had to appear to be "interleaved," so they covered about the same period of time. In order to verify this, the dates from each rater's chosen days were converted into numbers. The numbers were then compared with a Mann-Whitney-U test. If one rater's group of dates was significantly larger than the other (indicating more recent days), that twin pair was either omitted from the analysis or, if possible, the dates used were modified. The modification was to remove dates from one or both observers until there was not a significant difference in their dates. This process left 10 families in the analysis with an average of 33 days for both observers. The total number of words estimated for the children at the time of this analysis, going by the overall total provided by the journal for the entire study up until this period, was an average of 111.

Once the journal entries had been checked and modified if necessary, vocabulary lists were compiled (using the process described in the next section) for each observer and compared. On average, the two raters differed in the number of words they estimated for a child by 10.5 words or about .32 words per day. If this number is considered as a percentage of the average of 111 words the children had at the time, it amounts to about 10%, which seems rather high. However, the within-pair correlation of the ten pairs of listkeepers was .97 (for this correlation, the two raters within each pair of raters were ordered in ascending order of words they estimated; with random assortment the correlation is .94). Furthermore, (signed) co-twin differences in the number of words estimated during this period by the two raters were also highly correlated: $r = .89$. Thus the two observers appear to be in fairly good agreement both on the total number of unique words the twins said over the 66 day (average) overlapping period, and they were very similar in their estimates of the difference between co-twins. Since the difference between co-twin vocabularies will be the major dependent measure of this chapter, this fact is very important. It means there is apparently not a great deal of observer difference in estimating the difference between co-twins, and thus the shared environment will not be greatly inflated by this particular factor. However, the fact that observer estimates vary by as much as 10% is troubling and indicates that even though rater bias is small, it does exist and it merits further investigation.

2.3. Data extraction and some residual issues

As noted above, parents sent their journals in each week. The words reported in the journals were entered into Microsoft Excel spreadsheets. The words were generally entered as the adult word that the child intended to say according to his parents. For instance, the child may have said "ba" for bottle, but it was entered as bottle. In some cases, the parent's explanation of the word did not match the child's output: for instance, "choo-choo" would be recorded as the word but "train" would be written in the comments section as an explanation. In these cases, the word the child learned was assumed to be choo choo or choo choo train, not train. Such examples are common with animals sounds and names as well as with a few vehicles like train.

Different forms of the same word were combined into just one vocabulary item so that vocabulary estimates from this study would be comparable with those from other studies. For instance, in the CDI, each noun is listed in its singular. If we were to count both a singular and a plural as separate items, the child's vocabulary estimate would be larger by our estimate than by the CDI. Therefore, all inflectional variations of a word were collapsed into one word, as were different versions of a word such as mama and mommy. Words related by derivational morphology, such as an adjective and a noun made of the same root, were not combined (of course the parent had to indicate that the two words were different in order for us to know that).

Three problems arise from long-term diary data because of the developmental level of the children when their parents joined or because of lapses in record-keeping. The first is the problem of "late-starters"—twins who began the study weeks or months after their first words. When their first weeks' journals came in, they already contained between 10 and 50 different words. The second is the problem of "early-starters"—twins who began the study with just one or two words but failed to add any more for several months. With these early-starting cases, it was suspected that the first 1-2 words reported might be remnants of babbling that were interpreted as words by anxious parents. The third is the problem of "missing data," where journal-keepers took long breaks in the middle of otherwise good data. All of these groups can distort the results of this study not only by making the numbers for a child inaccurate, but also by making twin pairs appear more similar than they really are. For instance, if a pair of twins joins the study after they have said their first words, but we as experimenters think that the first words the parent writes down on the first day of the journal are really the first words for both twins, we will think they have said their first words on the same day. In fact, those twins may have said those words for the first time on a much earlier day, and not necessarily on the same day as one another. These kinds of effects will increase apparent twin similarity.

Note that it is in principle perfectly reasonable to use late and early starters in the data analysis, since our two interests—their age at vocabulary milestones and their overall rate of word learning—will be unaffected. For age at milestones, twins who did not reach a milestone during the study (either because it happened before they joined or after they left) simply will not be included in the analysis for that milestone. For rate of word learning, the meter can start and end at any point; we need not capture the first or 100^{th} words specifically to calculate rate. We just have to identify the correct beginning and end points. It is also in principle fine to use journals with missing entries, as long as we allow a catching up period after a long period of missing records before we start trusting the words we see as new. To find the correct starting and ending points, the following methods were used.

To deal with missing data, we had to determine how long a catching up period to allow before counting journal entries as new words. We did this by studying twins who were clearly not late-starters—those who started with just a couple of words and added them slowly. First, the point at which they reached some milestone (say, 20 or 30 words) was found. Then their journals were re-started on that week and the number of days it took them to catch up to the number of new words that had before their early data were thrown away was measured. For vocabularies under 20 words, it took about a week, and for greater than 20 words, two weeks. Therefore, for late-starters with less than 20 words, the first week of journal entries were not counted as new words, and for late-started with more than 20 words, the first two weeks of journal entries were not counted as new words. In total, 35 pairs of twins had some adjustment made due to a late start (17 DZ, 18 MZ). Of these 35, 16 pairs had 10 or fewer words after the first or second week (8 DZ, 8 MZ); 7 pairs had between 11 and 20 words (4 DZ, 3 MZ); 5 pairs had between 21 and 30 (2 DZ, 3 MZ); 3 pairs had between 31 and 40 words (1 DZ, 2 MZ); 2 pairs had between 41 and 50 words (1 DZ, 1 MZ); 1 pair had between 51 and 60 words (MZ); and 1 pair had between 71 and 80 words (DZ).

Early-starters were handled less strictly. In each case, a point where words were being added with some regularity (at least one per week) for at least one co-twin was chosen by hand before counting new words began for both twins. Nine pairs of twins (4 DZ, 5 MZ) were judged to be early-starters and had anywhere from a few days to a few weeks removed from the beginning of their journal data. In all cases, both twins were started at the same point in time.

Finally, to deal with large chunks of missing data, journal entries were not counted as new words for 1-2 weeks after a large gap (depending on how many words the twins had before the gap—again, the rule was 1 week for 20 or fewer words, 2 weeks for more than 20 words). In these cases, as with the late-starters, milestones of interest sometimes had to be thrown out, but their accuracy was questionable anyway.

Once words had been entered for all the journals a parent sent in and adjustments were made for early- and late-starters and missing data, a Visual Basic macro was written to detect what date a word was first said on. The result was a list of dates and the number of new words for each date for each child which served as the dependent measure on which most of the following analyses are based.

2.4 Subjects

In the vocabulary study, data will be reported for 43 pairs of MZ and 33 pairs of (same-sex) DZ twins. These and other subjects (some of whom dropped out, some of whom are still participating) were recruited through several methods. The majority of subjects came from Mothers of Twins Clubs, support groups for parents of twins. In addition, after the first year of the study, we contacted the National Organization of Mothers of Twins Clubs who, for a small fee, distributed flyers to hundreds of clubs across the United States and Canada. The second largest source of subjects was the internet, either via announcements on the Twins Digest mailing list or visits to our website (*web.mit.edu/jganger/Public/ourhome.html*). Some subjects were also obtained by newspaper ads, personal contacts at the 1995 Twins Festival in Twinsburg, Ohio, and word of mouth.

Subjects who expressed interest were interviewed over the telephone at our expense. Potential subjects were asked a battery of questions, and the study and its compensation schedule were explained (see Appendix A for script followed during phone interviews). The main purpose of the interview was to screen subjects. With rare exception, twins were not asked to participate in the study if any of the following was true:

- one or both had a history of chronic ear infections, defined as more than 3 infections in their lifetime or the use of prophylactic antibiotics prescribed by a doctor to prevent further infections.
- the twins were exposed to a language other than English on a regular basis.
- there was a history of speech or language disorders in the immediate family.

In addition, a combination of gestational age, birthweight, and complications after birth was used to determine if the twins were eligible. In general, 34 or fewer weeks gestational age was a flag and resulted in non-inclusion if birthweight was less than 4 lb. or if a prolonged hospital stay was necessitated by severe complications, such as immature lung development or apnea. It should be noted, however, that 4 pairs of twins who were included in the data to be presented below should not have qualified for the study because of their gestational age or birthweight; these pairs were recruited before the cut-offs described above were established. As it turned out, though, their language development did not appear to be different from that of the average twin in the study. (Their data will be reviewed separately in section 2.2.5.)

To diagnose zygosity, a questionnaire designed by Goldsmith (1991) was used. The questionnaire was mailed to parents and they filled it out on their own and returned it to us. Research by Spitz *et al.* (1996) has suggested that this questionnaire diagnoses zygosity with high accuracy when compared to DNA testing. In addition to the questionnaire, information on blood type and placenta type were also requested. Having different blood types is a definitive sign that a pair of twins are DZ, and having a monochorionic placenta is a definitive sign they are MZ. If this biological information was ambiguous and the four key questions provided ambiguous or conflicting information, or if the parent requested it, a DNA test was ordered to test zygosity. This test was done through a company called Affiliated Genetics using cheek-cell swabs sampled by parents and sent to the company through the mail. Our lab received a copy of the results sent to the parents.

The subjects analyzed in this study participated an average of 27 weeks and had an average of 105 words when they stopped. The participants were primarily Caucasian, came from all over the United States and Canada, and were an average of 64 weeks old (15 months) when they started keeping journals (recall that some of them already had some words when they began). Thirty-four pairs were male and 42 female. All were same-sex.

Several biological and social factors require special attention because of their known or suspected relationship with language development. Theses are gender, birthweight, gestational age, SES, and number of siblings. It is important to note any differences in these factors between these twins and the normal population, or between MZ and DZ twins.

In this study, the average gestational age was 36.9 weeks. Although 40 weeks is considered full-term for a singleton pregnancy, 36 is considered normal for a twin pregnancy (Noble 1980) so, on average, these babies were not pre-term. Gestational ages were similar for MZ and DZ twins (see table 2.4). The average birthweight was 5.63 pounds, but the MZ average was slightly lower than the DZ average (5.4 versus 5.9 pounds), a difference which was significant ($t = 2.77$, $p = .006$ 2-tailed). I will investigate whether this difference had any ramifications for rate of word-learning in the next section. Fortunately, though, the birthweight difference between co-twins was similar for MZs and DZs (.67 versus .75, not significantly different by t-test), allaying some of the problems specific to MZ placentation raised in the introduction. Typically, MZ twins have larger birthweight difference than DZ twins (Mittler 1969), but the trend was reversed or negated in this sample. That reversal is presumably due to our screening procedures: twins were not accepted in the study if one or both had a birthweight of less than 4 pounds, and MZs with damaging twin-to-twin transfusion usually have one member with very low birthweight.

One important difference between MZ and DZ twins was gender. 51% of MZ twins were female, whereas only 36% of DZ twins are female. Girls are known to mature faster linguistically than boys, and if such a difference somehow results in different variances, the heritability of language development could be different in boys and girls. Since more MZs than DZs are girls, any zygosity effects could be driven by sex differences. I will consider this possibility in section 2.6.

SES was proxied by father's education[9], which was coded as follows.

0 = less than high school

1 = high school completed

2 = some college

3 = college completed

4 = any post-graduate education

Father's education (and mother's, too) was significantly higher for DZs than for MZs. This effect is unsurprising since DZs tend to be born to older parents in general (parents' ages were not ascertained for this sample so this is only a speculation). The difference in level of education may also have been related to the fact that about one-third of the DZ twins in this study were conceived with fertility treatments, which are very costly, while none of the MZ twins were. However, of the 18 DZ subjects who responded to the question of whether fertility treatments were used, there was not a positive correlation between educational level and use of fertility treatments ($r = -.21$); 11 responded "yes" and 7 "no." I will come back to the SES difference and its possible impact on word learning when considering the results.

These important variables (birthweight, gestational age, and SES) were somewhat intercorrelated. The average birthweight of a pair of twins and their gestational age were correlated with $r = .73$ ($p = .000$). This close relationship is not surprising since the longer a baby is *in utero*, the higher its expected birthweight. There are smaller but significant correlations between average birthweight and SES ($r = .28, p = .007$) and gestational age and SES ($r = .18, p = .060$), which are probably due to the better nutrition and medical care available to higher SES families.

Thirty-five pairs of twins had at least one older sibling. (None had younger siblings at the onset of the study and the birth of a younger sibling typically forced the parent to drop out of the study.) However, for the purposes of studies that consider whether a child is first-born, all of the twins should be counted as non-first-born since having a twin takes away parental resources as much as having a younger sibling does. They must share physical and psychological resources with this twin from the time of conception, so any affects of having a sibling should be apparent in all twins. All of these factors are broken down by zygosity in table 2.4.

---

[9]The means for level of education for fathers and mothers were virtually identical within zygosity, so father's education stands for both here since it is more traditionally used to figure SES.

Table 2.4

Basic information	MZ	DZ
Number of female pairs	22	12
Number of male pairs	21	21
Mean age (in weeks) at beginning of study	64	64
Mean weeks participation	26	29
Mean number words at end	94	119
Mean gestational age	36	37
Birthweight (lb.)	5.4	5.9
Co-twin difference in birthweights (lb.)	.67	.75
Mean father's education (see rating scale in text)	2.7	3.3
Number of pairs with at least one sibling	19	16

2.5 Results: vocabulary milestones and rates

2.5.1 Measures

There are many ways to look at long term data from 76 pairs of twins. In this initial analysis, two kinds of measures were used, each with its own merits. The first is a milestone based approach, comparing twins' ages at several levels of cumulative vocabulary, including the 10^{th} word, the 25^{th} word, the 50^{th} word, the 75^{th} word, the 100^{th} word, and the 125^{th} word. This approach allows us to check whether genetic effects change as vocabulary increases and it provides several measures of genetic effects (though they are probably not independent from one another). One disadvantage of a milestone approach, though, is that it is prone to error. If a parent is mistaken in keeping their journal very early on, those mistakes accrue for each milestone. A related problem is that a child may not say his n^{th} word on a particular day even though he knows it and is capable of saying it because the context does not arise for that word. Another disadvantage is that not all the twins in the study had data for all of the milestones, so each point has less than the full set of possible subjects and hence statistical tests have less power to detect effects than they could with the full set. An alternative approach which avoids these problems is to measure the rate of word learning over the whole period. This is simply the number of words learned by a child during the study divided by the total number of days in the study. Although this rate may collapse across interesting distinctions at different vocabulary levels, it has the virtues that it is less affected by random errors, it allows a wider time window for words to appear in, and it allows the consideration of data from all the twins in the study.

2.5.2 Population means

Before testing genetic effects or heritability, the mean ages at the vocabulary milestones and the mean rates of word learning for MZ and DZ twins were inspected to ensure that MZ and DZ twins represent the same population of language learners on both these measures. The milestone data are summarized in Table 2.5, which shows the mean ages (in days and months) when MZ and DZ twins reached each milestone, the standard deviations of those means, the number of subjects at each milestone, and the result of *t*-tests comparing MZ and DZ twins. Only at the 125^{th} word were there significant mean differences between MZ and DZ twins, with DZ twins reaching the milestone at an earlier age. Since at every other point, comparisons indicate that MZ and DZ twins come from the same population, and since the 125 word milestone has the fewest subjects, I will assume this finding is meaningless.

Table 2.5

Milestone	Overall Age: mean (s.d.) in days and months	MZ Age	DZ Age	*t*-test for difference[10]
10th word:	466.0 days (78.0)	468.8 (88.1)	462.1 (61.9)	$t = .43$
	15.5 mos (2.6)	15.6 (2.9)	15.4 (2.1)	$p = .668$
N of children	106[11]	62	44	
25th	534.0 (89.1)	529.8 (83.7)	540.5 (97.4)	$t = -.64$
	17.8 (3.0)	17.7 (2.8)	18.0 (3.2)	$p = .526$
N	116	70	46	
50th	588.1 (93.7)	583.8 (74.0)	595.2 (120.4)	$t = -.60$
	19.6 (3.1)	19.1 (2.5)	19.8 (4.0)	$p = .552$
N	104	65	39	
75th	616.4 (97.3)	624.8 (66.9)	607.0 (122.8)	$t = .81$
	20.6 (3.2)	20.8 (2.2)	20.2 (4.1)	$p = .418$
N	80	42	38	
100th	640.2 (109.7)	649.7 (72.6)	631.7 (135.0)	$t = .69$
	21.3 (3.7)	21.7 (2.4)	21.1 (4.5)	$p = .492$
N	72	34	38	
125th	635.4 (77.8)	659.4 (69.1)	613.3 (80.0)	$t = 2.17$
	21.2 (2.6)	22.0 (2.3)	20.4 (2.7)	$p = .035$
N	50	24	26	

[10]All *p*-values in this table are 2-tailed, since a priori we have no reason to think that one twin type will reach a milestone at an earlier age than the other.
[11]This number is more than twice the number of MZ pairs because some children were included here whose twins did not reach their tenth word during the study

For the rate data, 152 children (86 MZ, 66 DZ) were available for the analysis. The mean rate (words/day) was .66 (s.d. = .75) for MZ children and .77 (s.d. = .78) for DZ children. These were not significantly different ($t = .88$, $p = .379$ 2-tailed), indicating that the MZ and DZ children represent the same population with respect to rate of word learning.

2.5.3 Genetic effects

Next we can turn to genetic effects—that is, MZ-DZ differences—in these dependent measures. For the milestone data, co-twins' ages were compared at the milestone of interest (say, 10 words) in order to find the difference between the two twins' ages. That difference score was used as the dependent measure. For the rate data, the differences between co-twins' rates were used as the dependent measure. The differences between MZ and DZ difference scores in both cases were checked using simple *t*-tests. If there are genetic effects, difference scores should be smaller for MZ twins than for DZ twins (note that this the opposite of what we expect to find with correlations, where MZs should have a higher number than DZs).

The average age differences for all twins at each milestone, as well as for MZ twins and DZ twins separately, are shown in Table 2.6. The data in that table show that at each milestone, the difference is in the predicted direction. That is, at each level of vocabulary, MZ twins have smaller age differences than DZ twins. The final column shows *t*-tests for MZ-DZ differences at each vocabulary level. The difference is significant at the $p < .05$ level at the 50^{th}, 75^{th}, 100^{th}, and 125^{th} vocabulary milestones, but not at the 10^{th} or 25^{th}. Bar charts, with error bars equal to ± 1 standard error, for each of these milestones are shown in Figures 2.1-2.6 at the end of this chapter.

Table 2.6

Difference in days between co-twins at all vocabulary milestones

Milestone	Overall difference	MZ difference	DZ difference	*t*-test for difference
10th word	16 (17.5)	13.5 (18.9)	19.3 (15.1)	$t = 1.18$
N of pairs	51	29	22	$p = .122$[12]
25th	14.7 (17.7)	11.7 (14.2)	19.4 (21.5)	$t = 1.64$
N of pairs	58	35	23	$p = .053$
50th	17.8 (19.1)	14.4 (13.2)	23.6 (25.6)	$t = 1.36$
N of pairs	51	32	19	$p = .049$
75th	19.6 (26)	12.5 (16.6)	27.4 (32.1)	$t = 1.88$
N of pairs	40	21	19	$p = .034$
100th	20.83 (27.5)	9.1 (9.2)	31.4 (33.9)	$t = 2.63$
N of pairs	36	17	19	$p = .007$
125th	19.7 (20.3)	12.0 (11.3)	26.9 (24.3)	$t = 1.93$
N of pairs	25	12	13	$p = .033$

This result is somewhat surprising on the surface, since we might expect the timing of a child's earliest words to be more driven by genetic factors. However, there are two possible explanations for the lack of significance at the early milestones. The first is that genetic effects increase over time. By the 50th word, genetic effects that were not present at the 10th word have come on line. This possibility is consistent with Reznick, Corley & Robinson's (1997) finding of increasing heritability for productive language between 14 and 24 months. The second explanation for increasing genetic effects is simply that at lower vocabulary levels, and hence at lower ages,

[12]All *p*-values in this table are 1-tailed, since we hypothesize in advance that the MZ differences will be smaller than the DZ differences.

there is less room to vary, and hence less room for MZ-DZ differences. Whatever the explanation, it is clear that there are genetic effects on individual differences at several milestones, and that effects are in the right direction to support genetic effects at all the milestones.

Turning to the data for rate, MZ twins had an average rate difference of .07 (s.d.= .06) words per day (n = 43 pairs) and DZ twins an average difference of .16 (s.d. = .20). words per day (n = 33). The MZ co-twin differences were significantly smaller on average than DZ co-twin differences ($t = 3.13$, $p = .002$). However, the MZ and DZ variances were also significantly different (by Levene's F-test for inequality of variances), making the use of the ordinary student's *t*-test questionable. To avoid the assumption equal variances of the standard *t*-test, two additional tests were performed. First, a *t*-test that does not assume equal variances was used. The difference was still significant ($t = 2.79$, df = 36.32, $p = .004$, 1-tailed). Second, a non-parametric test, the Mann-Whitney-U test, was also performed so that any differences in underlying distributions of the difference scores would not affect the outcome. This test also revealed a significant difference between MZ and DZ difference scores ($Z = 1.88$, $p = .030$, 1-tailed). The means and ± 1 standard error for the rate data are shown graphically in Figure 7.

For the first time, then, it has been shown that individual differences in early vocabulary learning have a genetic basis. This result does not mean that environmental factors are not important in determine differences in vocabulary development. Rather, it only means that when investigators search for the cause of individual differences in language development, they should look not only for environmental factors that differ but for genetic factors as well.

2.5.4 DeFries-Fulker analysis

Genetic effects, however, are not the same as heritability. Furthermore, the *t*-test analysis does not allow us to address environmental factors directly. Therefore, DeFries-Fulker (DF) regressions were performed on age for each milestone and for the rate data. Table 2.7 shows MZ and DZ correlation coefficients, as well as h^2, c^2, and e^2 for all the milestones. Table 2.8 shows the same for rate data.

Table 2.7 [13, 14]

Milestone	r_{MZ}	r_{DZ}	h^2	c^2	e^2
10th word	r = .967	r = .921	$h^2 = .09$	$c^2 = .88$	$e^2 = .03$ [15]
N of pairs	29	22	$t(47) = .45, p > .5$	$t = 4.88, p < .001$	
25th	r = .976	r = .956	$h^2 = .04$	$c^2 = .94$	$e^2 = .02$
N of pairs	35	23	$t(54) = .29, p > .5$	$t = 8.37, p < .001$	
50th	r = .966	r = .959	$h^2 = .01$	$c^2 = .94$	$e^2 = .05$
N of pairs	32	19	$t(47) = .06, p > .5$	$t = 7.83, p < .001$	
75th	r = .952	r = .941	$h^2 = .02$	$c^2 = .93$	$e^2 = .05$
			SE = .60	.49	
N of pairs	21	19	$t(36) = .08, p > .5$	$t = 5.65, p < .001$	
100th	r = .984	r = .941	$h^2 = .09$	$c^2 = .90$	$e^2 = .01$
			SE = .63	.50	
N of pairs	17	19	$t(32) = .33, p > .5$	$t = 5.29, p < .001$	
125th	r = .971	r = .897	$h^2 = .15$	$c^2 = .82$	$e^2 = .03$
N of pairs	12	13	SE = .70	SE = .60	
			$t(21) = .45, p > .5$	$t = 3.28, p < .01$	

Table 2.8[16]

DF analysis of rate of word learning

[13] All the values of h^2, c^2, and e^2 in this table are positive.
[14] The *t*-statistics in this table were calculated by dividing each statistic (h^2 or c^2) by its corrected standard error (see footnote 1), with df = N – k – 1, which is, for example, in the first cell, 51 (pairs) minus 3 (regressors) minus 1.
[15] No standard error estimates were available for e^2 because this is calculated indirectly from h^2 and c^2, rather than directly from the initial regression.
[16] All the values of h^2, c^2, and e^2 in this table are positive.

r_{MZ}	r_{DZ}	h^2	c^2	e^2
.99	.95	.09	.90	.01
n = 43 pairs	n = 33 pairs	$t(72) = .82, p < .4$	$t(72) = 3.91, p < .001$	

Despite the strong MZ-DZ difference in similarity we saw in the previous section, h^2 was quite small at each milestone, reaching its highest at .15 for the 125th word. At no point was it significantly different from zero, but with relatively small sample sizes and small effect sizes, that is to be expected (see Chapter 1). Interestingly, there appears to be a trend of increasing heritability as the number of words increases, but this is not a trend we can test yet. The magnitude of h^2 was similar in the rate data, at just under 10% and not significantly different from zero.

A more striking finding in both the milestone and rate data was large shared environment (c^2). c^2 ranged from .82 to .94 for the milestones and was significantly greater than zero at all points. c^2 was in a similar range for the rate data at .90. The implication of this finding is that individual differences in vocabulary learning are controlled largely by environmental differences and only to a small extent if at all by heritable ones.

However, before we take this finding at face value, it should be noted that the MZ and DZ correlations were both extremely high. In the dozens of twin studies on all different topics reviewed in Plomin, DeFries & McClearn (1990), there are none with such a high DZ correlation and few with such high MZ correlations. While it is possible that those correlations are correct, it seems quite likely that they were driven up by some feature of the method used in this study. Perhaps the subjectivity of figuring out what words infants are using allows too much room for rater bias, driving up MZ and DZ correlations. This possibility deserves more investigation than I have been able to give it in this thesis but remains a project for future research.

For now, though I will tentatively take these results as they appear and assume that, despite the presence of some genetic influence, there is very little heritability and large shared environment in vocabulary development. One can imagine, after all, that differences in word learning could be highly dependent on characteristics of the environment. Children can only learn the words they are exposed to, and perhaps their rate of learning is limited to the number of words that they are exposed to. Vocabulary development may be the one area behavior genetics has discovered for which shared environment plays an important role.

2.5.5 Eliminating nuisance variables

At this point, it is important to consider whether any of the extraneous variables that were raised in the preceding sections (MZ-DZ differences in sex, birthweight, gestational age, and SES) can account for the genetic effects we found. In addition, the four pairs of twins who did not meet the established cut-offs for birthweight and gestational age will be compared to the rest of the twin sample in terms of their language development.

Recall that male twins make up a larger proportion of the DZ group than the MZ group (63% versus 49%). Males are generally slower to mature in many respects, including language development, and it is conceivable that this delay relative to females causes males to have different variance and possibly different heritability for language development. If genetic effects in language are more potent in males than in females, sex differences could be driving the results instead of zygosity differences. To check this possibility, an ANOVA was performed with rate differences in word learning (as described above) as the dependent variable and both zygosity and sex as independent variables. Consistent with the results above, there was a large main effect of zygosity ($F = 8.47$, $p = .005$), but there was no effect of sex ($F = .09$, $p = .770$) and no zygosity by sex interaction ($F = .01$, $p = .917$). Thus we can be confident that the main genetic result of section is not an artifact of sex differences.

Birthweight and gestational age are also important variables that could affect the rate of language learning. At the extremes, low birthweight and gestational age cause pervasive maturational delays throughout infancy (though cognitive delays typically dissipate later in childhood; Wilson 1985). Both these variables were significantly higher for DZs than MZs. To check for possible effects of these variables on genetic effects in language, a multiple regression performed with the rate differences data, since using rate differences affords us the largest sample size available. Rate differences were used as the dependent measure, and zygosity (coded for MZ versus DZ) was an independent variable. Three nuisance variables were also entered as independent variables: birthweight (the average for each twin pair), absolute birthweight difference, and gestational age. The regression revealed that even with the nuisance variables partialed out, zygosity was still a significant predictor of magnitude of rate differences. Furthermore, none of the other variables were significant predictors on their own. The results of that regression are presented in Table 2.9

Table 2.9

Results of regression partialling out effects of several nuisance variables from zygosity on rate differences

Variable	**B**	**t**	**p**
Average BW	.02	.71	.481
Absolute BW difference	.02	.71	.482
Gestational age	.00	.01	.989
Zygosity	.09	2.67	.009

A different kind of problem is the possibility of differences between (this sample of) twins and the rest of the population. As we saw in Chapter 1, twins are suspected to be delayed in learning language and may be below average in other cognitive areas as well. However, there is no direct way to compare twins on language learning, since the methodology used here has never been used with a large population of singletons. Instead, the norms of the CDI will be used for comparison purposes, since the CDI and the journals provide similar vocabulary estimates (see section 2.2.2 and Reznick & Goldfield 1994).

Compared to the norms provided by the CDI manual, at all the milestones, both girls and boys were 2-4 months behind the median age at which the milestone should be reached. At the

10th, 25th, and 50th, the twins never get above the 25th percentile. At the 75th, 100th, and 125th, they started to catch up, reaching the 30th to 45th percentiles. Table 2.10 below shows how they compare. Although we must bear in mind that the journal records may not be perfectly comparable with the CDI norms, it seems that twins are living up to their reputation of being delayed in language development. However, they are only a few months behind, not several years as once feared. These subjects will require follow-up to ensure that they are do not develop speech or reading disorders as older children, but they appear to be developing only a little bit behind average, especially considering they were born an average of 3-4 weeks earlier than the population they are being compared to.

Table 2.10

	girls age and percentile	**boys age and percentile**	**age of median (50th percentile)**
10th	15 mo	16 mo	12-13 mo (girls)
	25-30th	15-20th	13-14 mo (boys)
25th	17 mo	18 mo	14-15 mo (girls)
	10-15th	15th	15-16 mo (boys)
50th	19 mo	20 mo	16 mo (girls)
	15-20th	20th	16-17 mo (boys)
75th	20 mo	21 mo	16-17 mo (girls)
	20-25th	20-25th	18 mo (boys)
100th	21 mo	22 mo	17-18 mo (girls)
	30th	35th	19-20 mo (boys)
125th	21 mo	21 mo	18-19 mo (girls)
	30th	45th	20-21 mo (boys)

Finally, the four pairs of twins who were recruited for this study before certain criteria for birthweight and gestational age cut-offs were established should be considered separately to see if their vocabulary development is different from that of the other twins. Of these four pairs of twins, 3 were MZ and 1 DZ. Their birthweights (expressed as an average for the pair) were 2.88 lbs, 3.50 lbs, 3.50 lbs, and 3.91 lbs, and their gestational ages were 29 weeks, 35 weeks, 30 weeks, and 32 weeks. However, despite being the smallest of the twins on these variables, their milestones of vocabulary development did not appear to be affected. Taking the age at the 25^{th} word as an example, these four pairs of twins had an average age of 531.5 days, compared to the mean for the entire sample of 534.0. The figures seem quite comparable, and so for now I will assume it is not imprudent to include these 4 pairs of twins in the sample

2.5.6 Motor development

Aside from providing basic biological information, parents also filled out a questionnaire indicating when their twins had reached various basic motor milestones: rolling over, sitting, crawling, and walking for the first time. Not all subjects were able to provide information on these milestones, and some joined the study before the questionnaire was in use. However, there is information on these milestones for about 30% of the twins in the analyses above. This information can be used to see whether genetic effects in productive vocabulary are related to genetic effects in early motor development. Since vocabulary production is dependent not only on knowing a word but on being able to say it, we might expect some relationship to obtain.

The largest amount of data was available for rolling and crawling, so only these two motor milestones were used. They were applied to the subjects from the rate analysis, since doing so maximizes sample size. Thirty-three pairs of twins from the rate analysis provided quantitative data for the onset of rolling and crawling—that is, parents reported exact or approximate number of days separating co-twins on reaching the motor milestone.

The correlation between co-twin age difference at the onset of rolling and co-twin rate differences in word leaning was $r = .32$, but the same correlation for crawling and vocabulary rate was $r = -.32$. In other words, age at first rolling was (somewhat) positively correlated with rate of word learning, while age at first crawling was negatively correlated. Kim Saudino (personal communication) suggests that crawling may be influenced by weight: heavier babies seem not to crawl as early as lighter babies, perhaps because they have more bulk to move. Consistent with this hypothesis, weight differences at the onset of the study had a correlation of $r = -.41$ with difference in age at crawling, implying that heavier babies do indeed crawl later. The positive correlation between age differences in rolling and rate differences presumably reflects a small relationship between the motor maturation needed to roll oneself over and the motor aspect of productive vocabulary.

The question, though, is whether these motor variables can account for rate differences better than zygosity. At this point, it is hard to tell for sure. A multiple regression was performed using rate differences as the dependent variable, and zygosity, age difference at rolling, and age difference at crawling as independent variables. Although several of these variables had significant first-order correlations with rate differences, none (including zygosity) had significant regression coefficients. Since this analysis brought us down to just 33 pairs of twins, perhaps it is just too few subjects to bring any effects to significance.

2.5.7 Rate: correlations

One of the goals of doing a study of such detail was to catch small and perhaps temporary fluctuations in rate that might themselves be genetically influenced.

Another way to compare the rate of word learning between twins is to find correlations between the number of words in their vocabulary over time, and see whether these correlations over time are higher for MZs than for DZs. Since both twins will necessarily be monotonically increasing over time (since total vocabulary is a cumulative function), partial correlations were computed instead, partialling out age (this can be thought of as partialling out a purely linear component, which increases linearly). The remaining partial correlation is a measure of how similar twins are in their non-linear fluctuations in vocabulary over time. In order to make the numbers more stable, the twins were compared using 17-day periods (see next section for why this number was chosen). The analysis was restricted to twins who started before 30 words and continued in the study until at least 60 words, since it is in this region (and beyond) that deviation from a linear growth curve was expected (see next section). This restriction left only 15 MZ and 9 DZ twin pairs, so results will necessarily be more speculative than in previous sections.

The mean partial correlation was .93 for MZs and .89 for DZs. On average, the MZ correlations were marginally higher than the DZ correlations ($t = -1.0$, $p = .154$ 1-tailed) [17], hinting at some genetic effects, though they are not significant. The difference between the two average correlations was also not significant by Z-test ($Z = .88$) but the direction of the effect is promising and the addition of more subjects should clear up the picture.

2.6 Vocabulary spurt

We saw that overall rate is slightly heritable. What about differences in rate like the presence of a vocabulary spurt? The vocabulary spurt has remained, ever since its detailed consideration in research like Nelson (1973) and Dromi (1987), rather poorly defined. There is general agreement that the spurt is some kind of increase in the rate of word learning around 18 months of age, or around 50 words total vocabulary. Several different criteria were used to try test the existence and heritability of the vocabulary spurt.

The daily vocabulary data for the twins considered above were first divided into 17-day periods, so that figures would be comparable with the data reported in Goldfield & Reznick 1990, one of the most recent and detailed papers on the topic. Then the average number of new words learned before 30 words cumulative and the average after 30 words cumulative were compared for half the children (one randomly selected child from each pair of twins). In other words, for each child, an average of the number of new words per 17-day period was found for the period before 30 words and the period after 30 words. These numbers were calculated and used for all the children, even if they only had data below or only above 30 words. All of the "under 30" averages were grouped together as were the "over 30" averages, so that some children had a data point in both groups and some only in one. Combining data in this way violates independence, which is an assumption of most statistical tests based on two samples, but they were compared using *t*-tests nonetheless to get an idea of the difference.

Judging by the outcome, there is strong support for a spurt in general. For the periods under 30 words, children learned an average of 4.5 words per period, while for the periods above 30 words, they learned an average of 13.2 words. Though the variance was much greater for the period above 30 words (56.3 versus 6.0), the difference was significant by *t*-test ($t = -7.88$, $p =$ 1.8E-12; n = 51 below 30 words, n = 54 above). Although the *t*-test may not be accurate because the observations are not independent, it should be clear that this mean difference (4.5 versus 13.2) is substantial and supports the existence of a vocabulary spurt in a larger number of children than has ever been studied in such detail.

But what about genetic influence on the spurt? Since almost all the twins had what would be considered a spurt by any reasonable standard, it would be foolish to look for the heritability of the presence versus absence of a vocabulary spurt. Instead, we can test the strength of the spurt: maybe some children have a stronger growth spurt than others. To try and test this, the MZ and DZ twins were separated and then again compared on rate of word learning before and after the 30-day mark. For each child (in twin pairs with data before and after 30 words) I pitted the average number of words learned per 17-day period before 30 words against that average after 30 words for each child and compared them by *t*-test for each child. While the test itself is not valid because the observations from each child are not independent, the effect size can give us some point of comparison between twins. The correlation between *t*-statistics for MZ twins was .86 (13 pairs), while it was .24 for DZ twins (9 pairs), suggesting large heritability. To avoid the violation of independence problem, I also compared co-twins simply on the difference between their mean rate of learning new words before and after 30 words. In other words, each child had a difference score consisting of his rate after the 30-word mark minus his rate before the 30-word mark. These

---

[17] *r*'s were converted to *Z*'s using the Fisher transform before averaging and then converted back to *r*'s for

scores will be more affected by variability than the t-statistic but are worth considering since they carry no statistical assumptions. Using this measure, the MZ correlation was .92 and the DZ correlation .46. With only 22 pairs total, we will have to wait for more data before reaching any conclusions. However, the strong difference in the predicted direction is promising and suggests that differences among children in the strength of a vocabulary spurt may not be driven completely by environment factors such as parents' willingness to play the "naming game" (Goldfield & Reznick 1990); there may also be a genetic component to that variation.

2.7 Vocabulary content/categories

Aside from actual rate of word learning, we can also test the heritability of the content of early vocabularies. If some children are noun-learners while others are social learners, as Nelson (1973) suggested, perhaps the differences are due to genetic differences in interest, disposition, or even the language faculty itself.

The first test we can do is to see whether not just the categories but the actual words children learn are heritable. On the one hand, we might expect total overlap in the words twins learn (both MZ and DZ) because children can only learn the words they are exposed to. On the other hand, if DZ twins are more disposed to have different interests and different levels of attention, they may learn different words. To find out, the number of overlapping words was counted out of three milestones, 25 words, 50 words, and 100 words. It was also calculated as a percent of total number of words (whatever that number was) for each pair. For 25 words, MZs had an average of 17 overlapping words and DZs 15, and this difference, though small, was significant by t-test ($t = 2.31$, $p = .013$ 1-tailed; n = 33 MZ pairs, 22 DZ pairs). For 50 words, the difference was again significant (35 versus 32 words; $t = 1.93$, $p = .030$; n = 27 MZ pairs, 19 DZ pairs). For 100 words the difference was less robust (73 versus 69 with $t = 1.3$, $p = .101$; n = 15 MZ pairs, 17 DZ pairs), perhaps because there were fewer subjects at this point. Finally, as a

---

interpretation.

percentage of the total words learned by a pair of twins, MZ twins have 67.4% overlap while DZ twins have 60.8% overlap, a difference which is significant by *t*-test ($t = 2.32$, $p = .012$).

The correct interpretation of this finding is not obvious. If the actual words children learn are genetically influenced, even slightly so, that presumably reflects a shared interest in the same things or shared level of attention to the words that are used in the environment, or that shared cognitive abilities result in a similar way of categorizing the world which affects word learning. It is not conceivable that actual English words children learn are inherited. Another possibility is that the genetic influence effect on word overlap is a by-product of having a similar rate of word learning. Many words in a child's environment change with the seasons, with the holidays, and with developmental levels (e.g., weather terms, holiday items, appropriate clothing and activities). If twins are ready to learn their 25th (or whatever) word at different times, the words available to them may be different. So the difference in overlap could just reflect rate differences.

Another possibility is that, the increased overlap for MZ twins could reflect observer error. Perhaps, as was mentioned earlier, parents of MZ twins are more apt to get confused and write down the same words for their twins. The videotape data reviewed above suggest such confusion happens with low frequency, but we do not yet know whether it is greater for MZ than for DZ twins.

Now that we have looked at actual words, we can also take a look at the categories of words children learn. A typical way to do this (suggested by Nelson (1973)) is to divide words into nominals (names for things, except nominals) and non-nominals (everything else, including proper names). This is typically done as a proportion of the first 25 or 50 words. I will consider the number of nominals out of 25, 50 and 100 words and compare co-twins by calculating difference scores (difference in the number of nominals) at each of these levels.

Out of 25 words, the MZ and DZ twins had similar difference scores: 1.8 for MZs and 1.6 for DZs (n = 35 MZ pairs, 20 DZ pairs). Out of 50 words, DZs had slightly larger difference scores on average: 1.6 for MZ and 2.6 for DZs, significant by *t*-test ($t = -1.79$, $p = .027$ 1-tailed), and out of 100 words, the effect is even stronger, with a difference score of 2.6 for MZs and 4.6

for DZ, for a $t = -2.41$, $p = .011$. While this effect may also be a by-product of the word overlap effect, notice that this effect is stronger as the number of words increases, while the genetic effect of actual overlap decreased at 100 words, so they are not completely related.

In addition to counting the number of common nouns, the number of proper names in each child's lexicon were also counted as a proportion of 50 and 100 words and compared the number between twins. For 50 words total, the number of proper nouns had an MZ correlation of .88 and a DZ correlation of .50 (N = 30 MZ, 20 DZ). Difference scores between MZ and DZ twins were significantly different, with DZs having larger difference scores on average than MZs (DZ mean = 2.1, MZ mean = 1.1, $t = 2.36$, $p = .012$ 1-tailed). At 100 words, the MZ correlation was also larger than the DZ correlation but the average difference scores were not significantly larger for DZs than for MZs, perhaps because the sample size is smaller at 100 words (N = 17 MZ, 19 DZ) or perhaps because the effect is only present with a small vocabulary because category differences in proper names just tend to even out over time over all individuals.

These findings suggests that the differences among children in terms of the types of language learners they are, first pointed out by Nelson 1973, and generally assumed to be driven by environmental differences, could have at least a small genetic influence. Perhaps some children are actually predisposed to be labellers and other more social-language users due to underlying personality or cognitive differences. Studies like Nelson's which correlated maternal language and social behavior with the child's style of language learning clearly confounded environmental and genetic variance; children's language development is surely affected by the environment their parents provide, but these studies had no way of teasing apart how much of the parent-offspring correlation was due to environment and that which was due to heritability.

2.8 Discussion

For the first time, genetic effects have been established for very early vocabulary production in infants. Individual variation in the speed of learning words has a small heritable component (sections 2.5 and 2.6), and fluctuations in rate may also have heritable component (section 2.6.3). Furthermore, individual differences in vocabulary processes and properties that were formerly thought to be driven by environmental differences also have a genetic component. The vocabulary spurt, an increase in the rate of word learning between 30-50 words cumulative, appears to have a genetic component and perhaps more so than overall rate alone. The categories of early words—another process that was thought to be driven by parental word usage (Nelson 1973)—also appears to have a heritable component, though it is small. Thus, our first goal of establishing heritability in early word learning has been met and the second goal, estimating heritability, has also been attempted. However, as was noted throughout the text, all the heritability estimates should be interpreted cautiously. They are not reliably different from zero with current sample sizes, and in a few cases population differences may obtain between MZ and DZ twins that would invalidate their comparison.

A finding that is even more clear was a large contribution of shared environment, c^2, in every analysis. Children's environments apparently have a large influence in determining when they begin to talk and how quickly they acquire words. However, recall that shared environment may be inflated by two factors in this study: twin confusions in word attribution and observer (parent) bias of any kind in keeping diaries. Both these possibilities merit further investigation before c^2 estimates are taken seriously. However, the presence of c^2 in individual differences in language development has been firmly established.

The finding of relatively large c^2 and relatively small h^2 is similar to the model for early productive language in Reznick, Corley & Robinson 1997. Recall that at 14 months, which slightly before the time of the 10th word in the present study, h^2 was estimated at .01 (ns), c^2 at .35. At 20 months, which is equivalent to the 75 word milestone, Reznick *et al.* estimated $h^2 = .25$ (ns) and $c^2 = .49$. At 24 months, which is a few months past the last milestone in the present study, Reznick *et al.* report $h^2 = .38$ and $c^2 = .40$. At the earlier points of measurement (14 and 20 months), Reznick *et al.* found that heritability was not reliably different from zero but shared environment was. The results of this study are identical qualitatively, though the actual estimates are quite different.

In addition to establishing h^2 and c^2, the milestone data suggested a trend of increasing heritability over time. The last milestone, 125 words, showed the largest heritability of all. The trend of increasing heritability over time is typical in developmental studies and was also reported by Reznick *et al.*, who suggested that a new genetic effect begins at 20 months and continues at 24 months. However, this effect should be interpreted with caution h^2 at 125 words also had a large standard error (it had the smallest number of subjects). More work will have to be done before this trend can be validated.

Figure 2.1

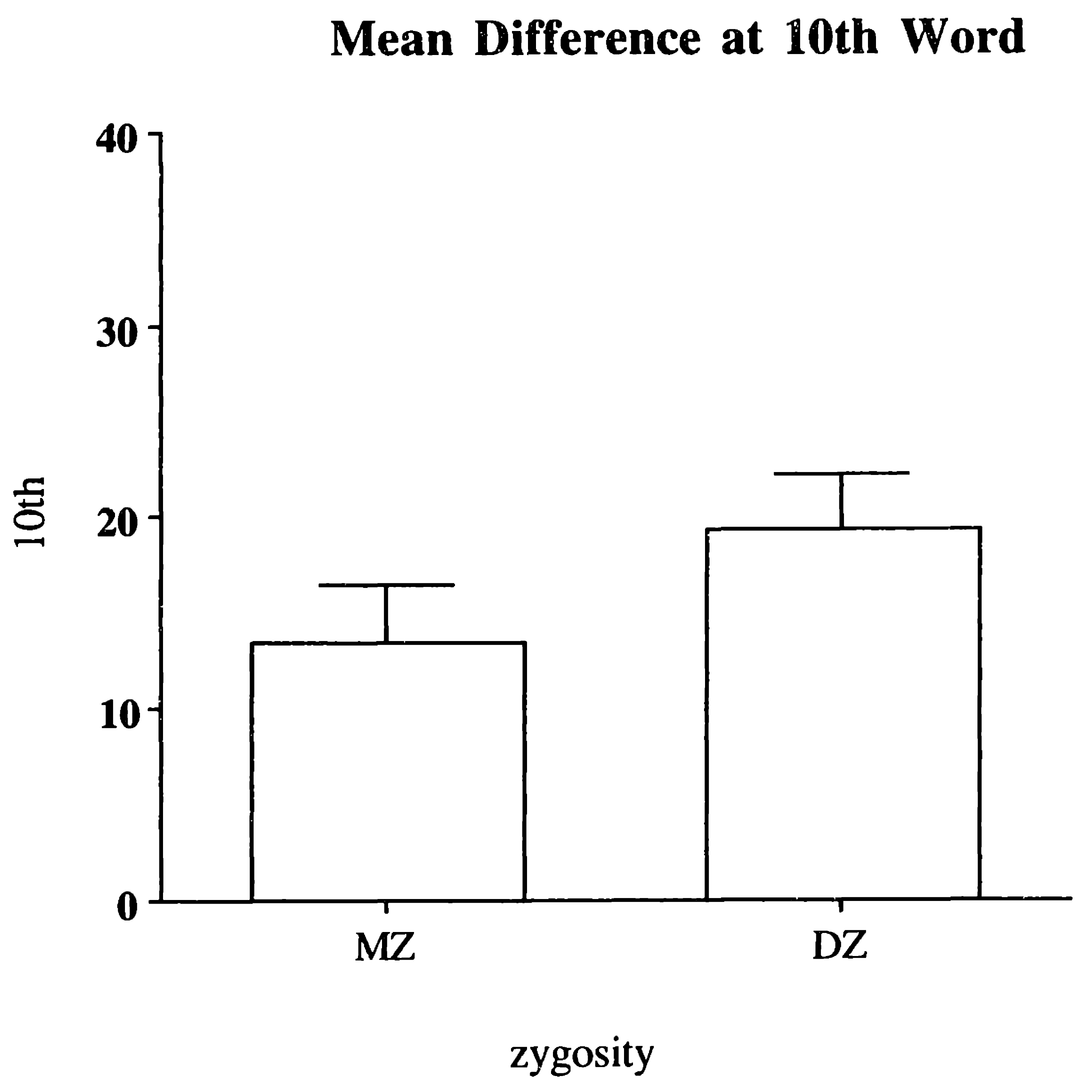


Figure 2.2

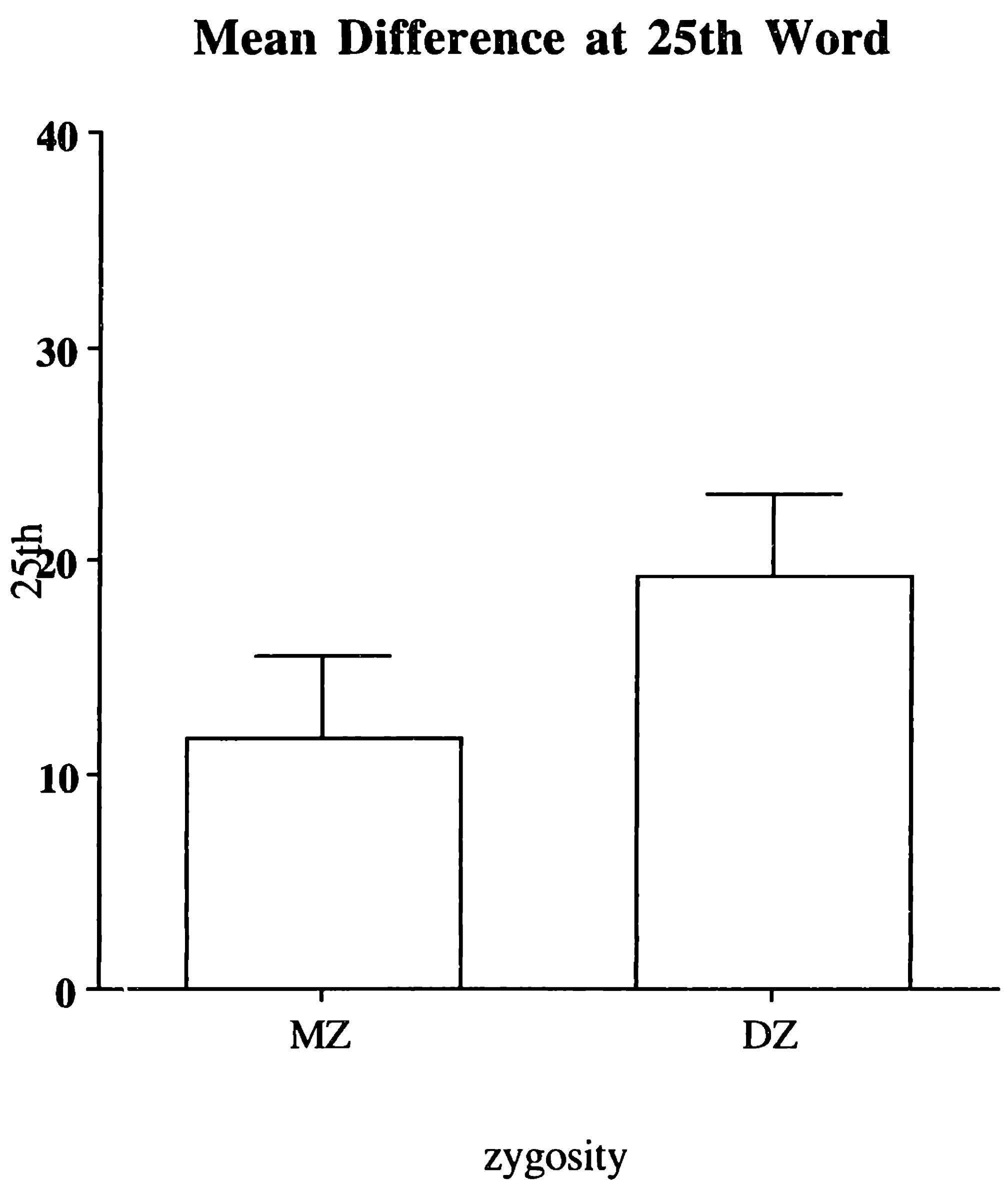


Figure 2.3

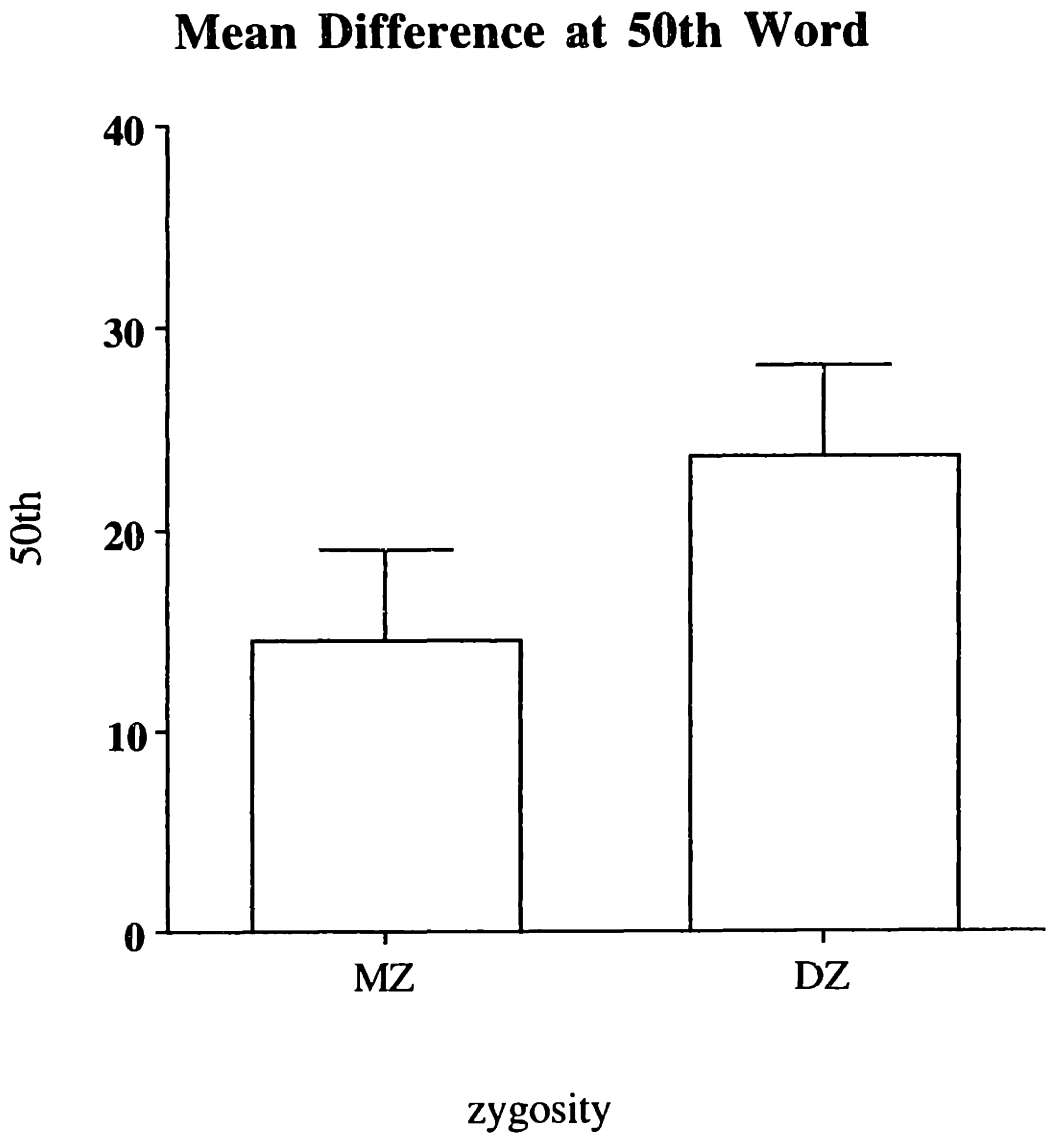


Figure 2.4

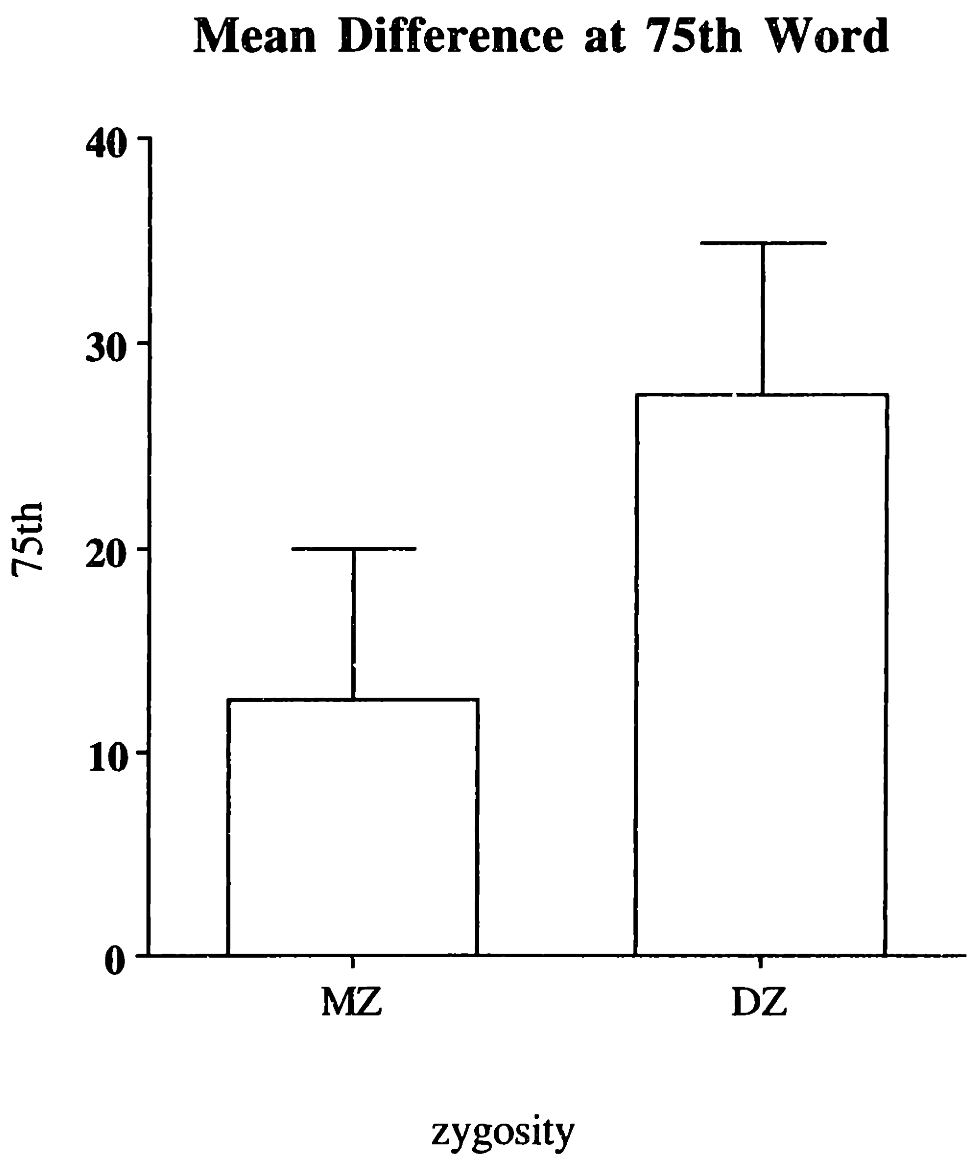


Figure 2.5

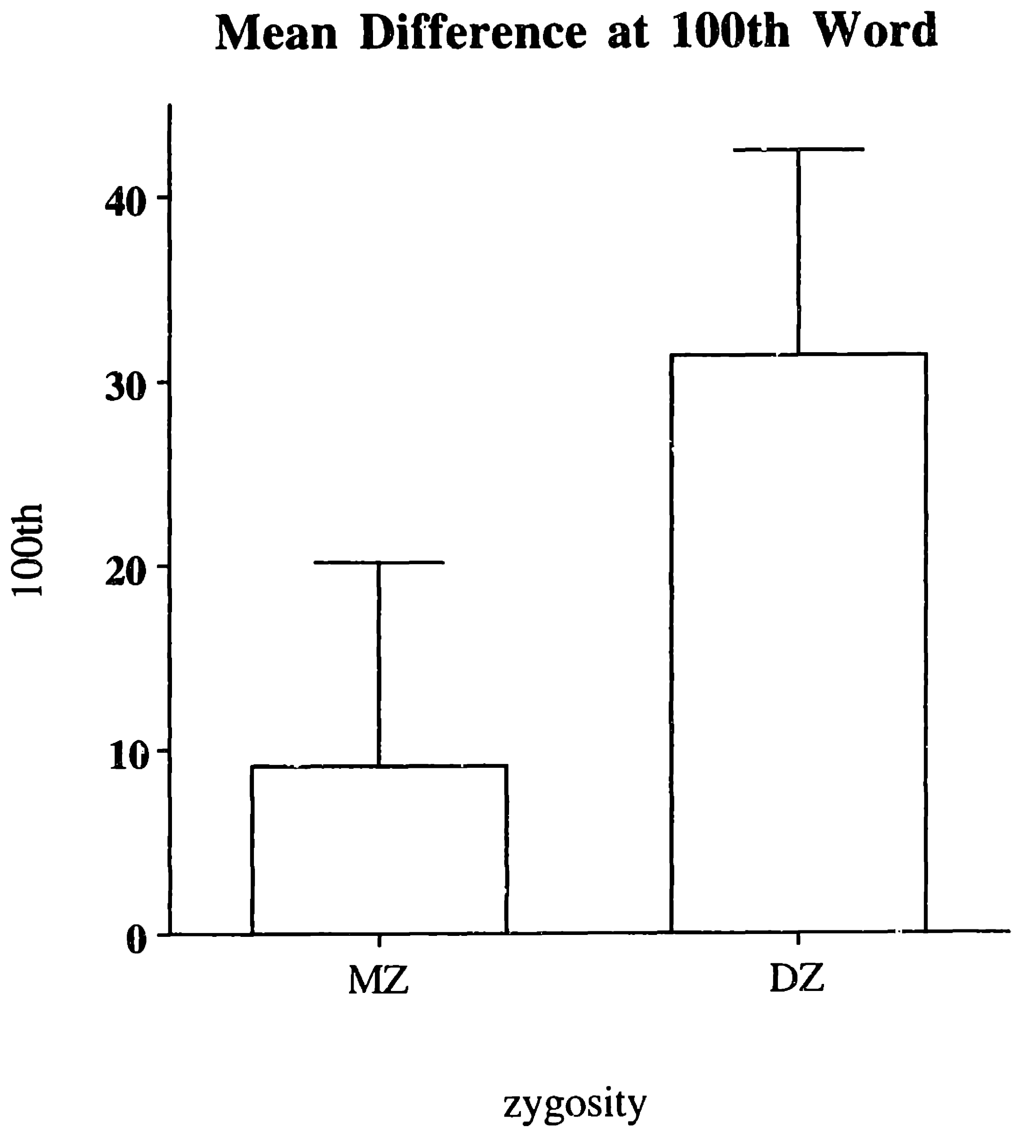
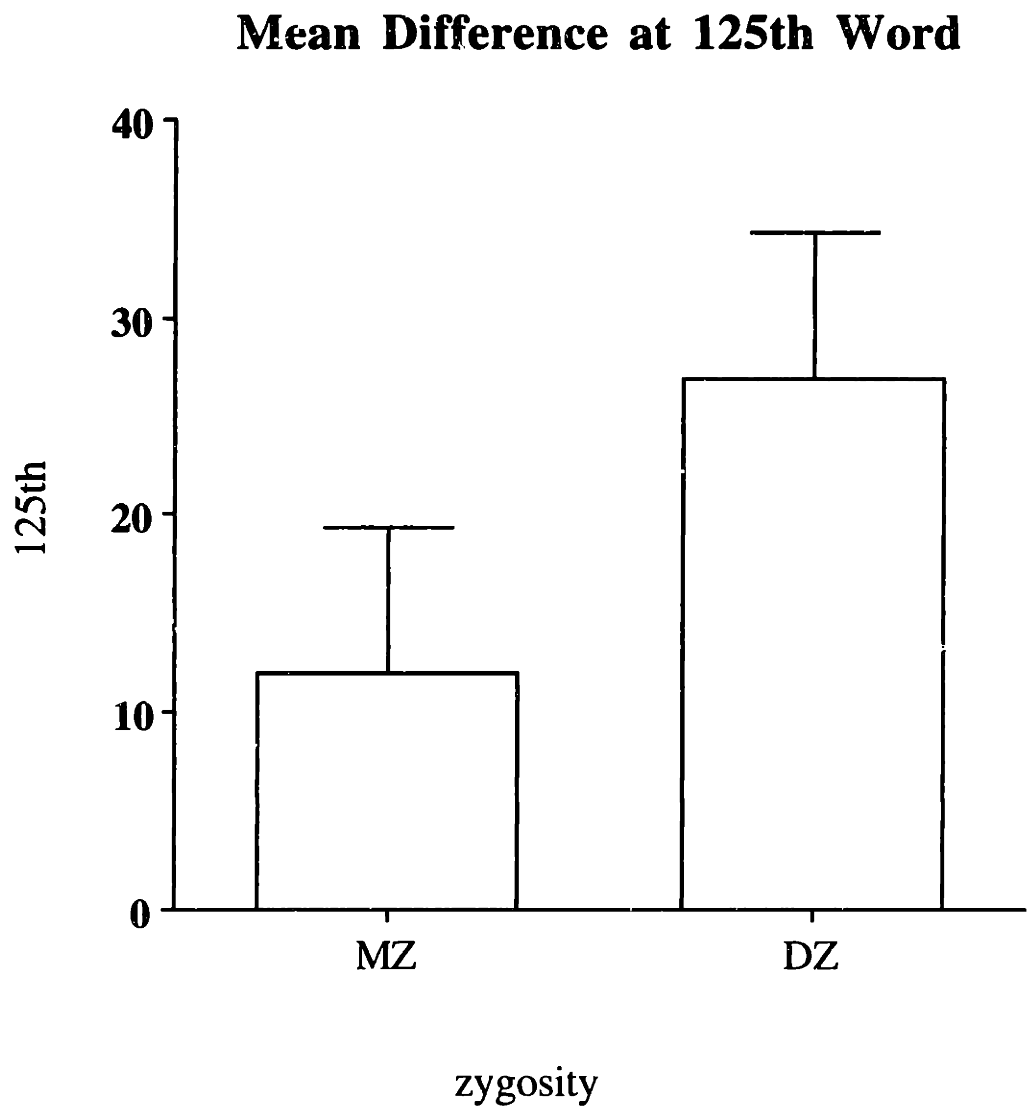
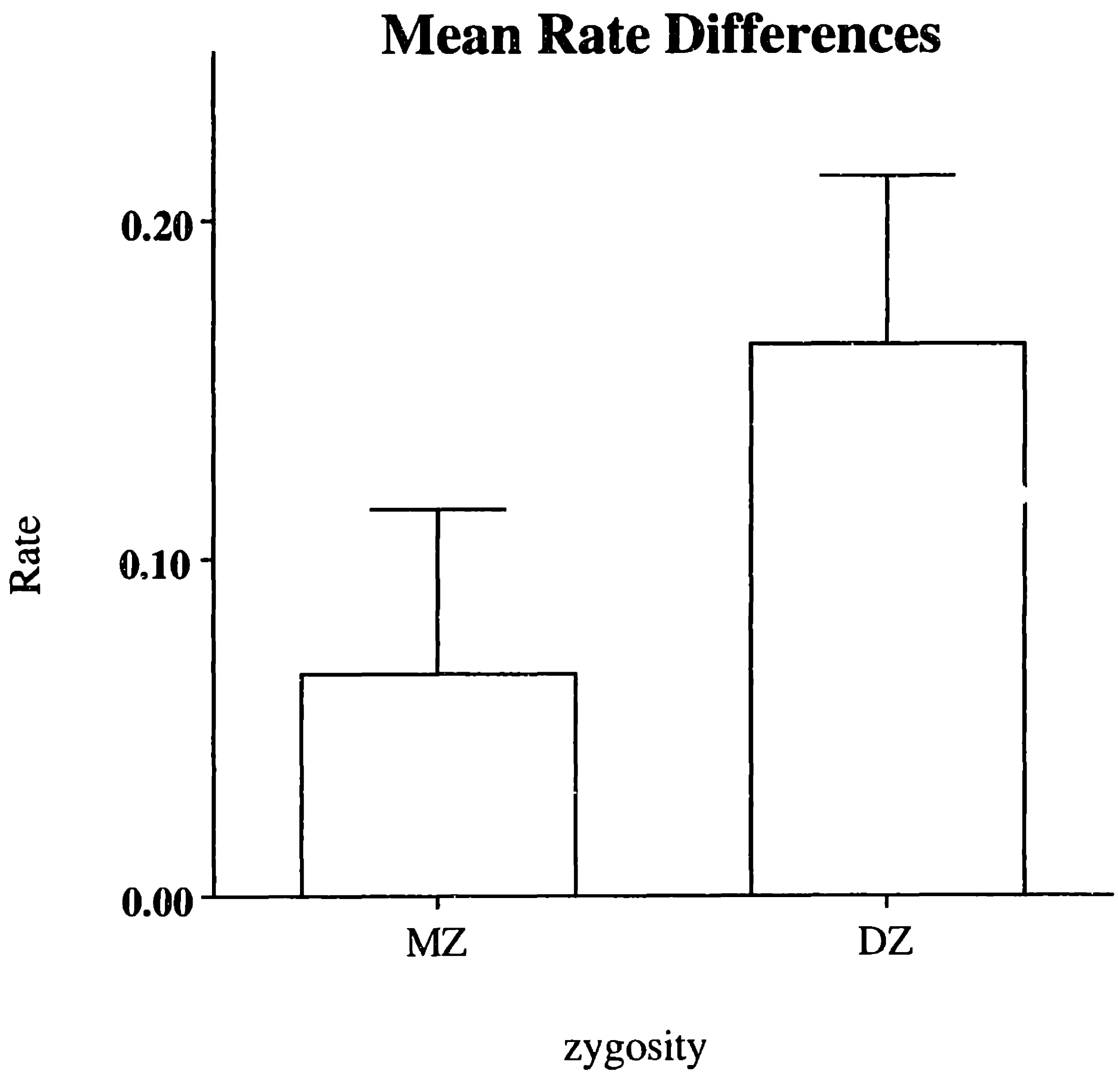

Mean Difference at 100th Word
40
30
20
10
0
100th
MZ
DZ
zygosity

Figure 2.6



Mean Difference at 125th Word
40
30
20
10
0
125th
MZ
DZ
zygosity

Figure 2.7



Chapter 3

First word combinations

3.1 Introduction

After children learn their first 50-100 words, the next milestone they approach is combining words, 2 or 3 at a time, into rudimentary sentences, such as "more juice" or "hi dada." To a normal adult, these combinations sound like crude and unsophisticated attempts at English. However, these combinations actually show considerable knowledge of the syntax of the child's parent language. For example, in English, children respect the rules of syntax in the only way they can at this stage: using word order. When a child wants more juice, he will say "more juice," not "juice more." And "bye-bye mama" means good-bye to mother, while "mama bye-bye" means mom is leaving (Brown 1973; Braine 1976).

Although much research has been focused on semantic and syntactic characterizations of the two-word combinations used by children in the earliest stages (e.g., Brown & Fraser 1963; Braine 1963; Bloom 1973; Bloom, Lightbown, & Hood 1976), the focus of the present study is on quantitative, not qualitative, variation among children. The interest of this study is in individual differences in the onset and rate of producing word combinations.

Multi-word phrases begin at a developmental level of about 50 words or an age of about 18 months, around the same time as the vocabulary spurt. However, just as with the spurt, there is a lot of variation in the timing of this milestone. McCarthy (1954), citing National Society for the Study of Education (1929), reports a range of ages from 1;3 to 2;0 in the onset of word combinations. Bates, Dale & Thal 1995, citing norms for the MacArthur CDI, state that by age 18 months, 11% of parents report that their children are already using some combinations and by age 25 months, only 19% do not yet report them. In other words, 70% of children begin combining words between 18 and 25 months, and the remaining 30% before or after that range. We are interested in finding out if any of that variation has a genetic basis.

No twin or other genetic study has yet addressed the issue of the heritability of this earliest of syntactic milestones. Even more than the use of single words in isolation, the process of combining words goes to the very heart of human language: its infinite combinatorial nature. It is hard to imagine how this process could be driven by the environment, since from their earliest productive phrases, children do not imitate what they've heard but produce novel combinations. Of course, the grammar of their parents' language must be induced from input, so some relatively rich linguistic environment is necessary for children to adequately converge on a grammar. But other than this very basic need, the onset of syntax should be largely driven by internal factors: overall mental maturity and/or specific linguistic maturity.

3.2 Methods

To track children's first phrases, a parent report technique that is basically identical to the one described in the previous chapter was used. Parents who were already participating in the first words (vocabulary) study of chapter 2 and whose children were saying a substantial number of words (e.g., 30-40) were asked to report phrases along with single words in their journals. In some cases, parents were instead recruited when their children already had an estimated vocabulary of 50-100 words but no word combinations. These parents were asked to write down only the phrases their twins used, not the individual words, and were given special instructions for this purpose. (See Appendix C for a copy of them).

The same advantages and disadvantages of a parent journal which were discussed in the last chapter presumably also apply here. Unfortunately, there are no supplementary data (like videotapes or CDI records) specifically based on multi-word combinations, so it is impossible to make the same kinds of comparisons that were made in Chapter 2. However, there is no reason to think that the issues are different here.

3.2.1 Data extraction

As with single words, a list of unique word combinations was extracted from the computerized journal entries using a simple Microsoft Excel Visual Basic macro, along with the dates on which each of these combinations occurred. Two or more combinations were collapsed into one if they consisted of the same words—or forms of the same words—in the same order. The same criteria for collapsing different forms of a word that were used in the vocabulary study (described in section 2.3) were applied here.

It is not obvious how one should measure the onset of word combinations. Since we are interested in using the first word combinations as an indicator of syntactic knowledge, it is not enough simply to ask parents whether their children have said any phrases. Instead, it is necessary to separate the truly productive combinations, where the child combines words into a novel phrase that has a meaning composed of its parts, from fixed or repeated phrases. Some obvious examples of fixed phrases might be "thank you," "good bye," or "all gone." Children often say such phrases weeks or months before they begin producing novel phrases of their own. However, discounting these phrases altogether as possible productive combinations would be too conservative, since these phrases really are made of two words for adults. To call "thank you" one word is tantamount to saying that the "you" in "thank you" has no relation to the word "you" in any other context. Presumably the child will learn that fact at some point, but we cannot know when that point is, at least not with naturalistic diary data. Therefore, to find the phrases that are truly productive, an objective criterion was applied to every phrase that was made of two or more words, even ostensibly fixed phrases like "thank you."

The criterion is inspired by that espoused by Pine and colleagues (Pine, Lieven & Rowland 1996; Lieven, Pine & Dresner-Barnes 1992) for "constructed" phrases but is unique in the literature as far as the author can ascertain. Our criterion was that, for a combination to qualify as productive, all words (or forms of the same words) used in that combination must have been used in previous combinations. For example, "more juice" would not qualify as productive until phrases using each of the words separately, such as "more milk" and "juice cold," had already been used. All multi-word combinations, even apparently fixed ones like "thank you," were thus afforded the opportunity to qualify as productive by this criterion. This criterion for productivity is a little more strict than that described by Pine and Lieven and their colleagues. Their diagnostic for a "constructed" or productive phrase was that it "contain one or more words or phrases [that had occurred] independently in the child's vocabulary, together with a word or phrase which [had] occurred in the same position in at least two other previous multi-word utterances." (Pine, Lieven & Rowland 1996, p. 579). While ours is a little different in detail, it is similar in spirit, easier to compute in practice (since position is not important), and captures more intuitively the combinatorial nature of grammar because all component words of a productive phrase must be veteran elements of combination.[18]

Word combinations were classified as to whether both words had previously occurred in (different) combinations or not (coding with 1 or 0) using a program written in the Perl programming language. This program examined all the phrases from one child and the dates on which they occurred and determined for each phrase whether a matching word occurred in a previous phrase. Before running the program, another program was used to check for common words which have multiple forms (e.g., *mama* and *mommy*) and morphological variations of the same word (e.g., plurals and possessives) and changed them all to the same form. Finally, the author or a research assistant checked the files by hand for other instances of alternative forms of the same word and any other inconsistencies. The output was then checked by hand for the earliest date that contained a phrase coded as productive.

3.3 Subjects

36 pairs of twins (19 MZ and 17 DZ) had enough data to be tested on the productivity criterion. That is, these pairs had enough phrases to meet the productivity criterion at least once. Of these, 13 MZ and 13 DZ pairs also participated in the vocabulary portion of the study described

[18]Aside from the criterion that all words in a combination be used in previous combinations, four other criteria were also tried: (1) that each word in the phrase be used previously in isolation; (2) that each word be used previously either in isolation or in another phrase; (3) that at least one word be used previously in isolation, and (4) that at least one word be used previously either in isolation or in a phrase. For each of these coding schemes, MZs reached the criterion at more similar ages than DZs. However, because of large variance, the effects were not always significant at the $p < .05$ level. The criterion used in the text seemed to provide the most stability (i.e., the smallest variance).

in Chapter 2, while the remainder were recruited specifically to track their word combinations. All subjects were recruited with the same methods described in chapter 2 and the same screening procedures were applied.

Table 3.1 below provides basic biological and SES information for the 36 pairs of twins whose data were used for the word combination study. Several of these factors were significantly different between MZ and DZ twins. MZ twins had significantly shorter gestational age ($t = 2.48$, $p = .009$ 2-tailed), significantly lower birthweights ($t = 2.73$, $p = .010$ 2-tailed), and significantly lower ratings on father's education ($t = 2.36$, $p = .024$, 2-tailed). However, as with the subjects in Chapter 2, there was no difference in co-twin birthweight discrepancies between MZ and DZ twins (and in both cases the mean difference is less than 1 pound), so we can probably rule out any complications secondary to MZ placentation (see Chapter 1). These MZ-DZ differences are similar to those reported in Chapter 2, which is not surprising since 26 of the 36 pairs were included in that analysis.

Table 3.1

Basic information	MZ	DZ
Number of female pairs	14	10
Number of male pairs	5	7
Mean gestational age	35.7	37.7
Mean birthweight (lbs)	5.09	5.93
Difference in birthweights (lbs)	0.73	0.73
Mean father's education (see rating scale in Chapter 2)	2.8	3.5

3.4 Results.

Three measures were used to test heritability. The first is milestone based. It is simply the age (or difference in co-twin ages) when the first productive word combination, as defined above, was used. The second is rate based. It is the number of multi-word combinations of all kinds produced per day. The third is also rate based, but instead of using the rate of producing word combinations in general, a rate was computed that consisted only of productive combinations.

3.4.1. Age

MZ twins uttered their first productive multi-word combinations at an average age of 21.2 months, DZs at an average of 19.3 months. DZs reached this milestone before MZs ($t = 2.38$, $p = .023$ 2-tailed). (The question of whether this difference affected MZ-DZ differences in co-twin difference scores will be addressed below.) As for difference scores, MZ co-twin difference scores were significantly smaller than those for DZs, with 22.7 days mean difference for MZs and 48.7 days mean difference for DZs; $t = 1.83$, $p = .038$, 1-tailed, suggesting genetic effects. Figure 3.1 shows bar graphs with error bars for this difference.

The MZ correlation for age at first productive word combination was .96, and the DZ correlation .33. Since the number of subjects in both the MZ and DZ correlations is small, scatter plots of age are shown for each (Figures 3.2 and 3.3). In both cases, there are no outliers nor is there any other obvious reason to think the correlations are spurious. Since the MZ correlation was more than twice the DZ correlation, heritability is technically greater than 1. However, it is not possible for h^2 to be greater than 1, so in order to constrain its value to lie between 0 and 1 in the DF analysis, c^2 was constrained to zero. With this constraint, h^2 was estimated at .90 ($t(33) = 6.92$, $p < .001$) and e^2 at .10.

3.4.2 Rate

Like vocabulary milestones, a milestone signifying the beginning of productive 2-word utterances can be skewed by errors in reporting or sampling. The overall rate of producing multi-word utterances may be more stable. This rate was defined simply as the number of multi-word utterances produced per day, productive or not. There is some danger that the rate of multi-word utterance per day may simply reflect talkativeness, and this issue will be discussed more thoroughly in a later section. However, since only unique phrases were counted each day, it must also reflect grammatical activity. A child who said the same phrases over and over in a day had a lower rate than one who used a several different phrases each day, so talkativeness is not the only factors controlling this rate.

Differences between MZ and DZ twins were examined by finding the difference between the rates within each pair of twins and using those difference scores in a t-test. 16 MZ and 18 DZ twins had enough data for this analysis (i.e., had more than one phrase). Once again, MZ twins had significantly smaller difference scores than DZ twins MZs had an average co-twin difference of .23 combinations per day and DZs .69 (t = 2.07, p = .011 1-tailed). Figure 3.4 shows a bar graph for this difference with error bars representing 1 standard error around the mean. The MZ and DZ correlations were .93 and .17, so once again c^2 was constrained to zero in the DF regression. The analysis revealed h^2 of .84 ($t(31)$ = 5.25, p < .001) and e^2 = .16.

The rate of generating *productive* multi-word combinations, as judged by the criterion described above, also showed genetic effects. Within twin pairs, the mean difference between MZs was significantly smaller than that for DZs—.12 versus .43 (t = 2.14, p = .02). Figure 3.5 shows a bar graph for this difference with error bars representing 1 standard error around the mean. The MZ correlation was .95, the DZ correlation .03, so once again c^2 was constrained to zero in the DF regression. h^2 was estimated at .81 ($t(31)$ = 5.06, p < .001) and e^2 at .19. Tables 3.2 through 3.4 below summarize all the word combination data.

Table 3.2

Group differences (MZ versus DZ) in age at first productive combination, rate of combinations, and rate of productive combinations

	MZ	DZ	t (p-value)[a]
mean age at 1st productive combo in months	21.2	19.3	2.38 (.023)
rate of new combos per day	.98	.99	-0.03 (ns)
rate of productive new combos per day	.39	.42	-0.20 (ns)

[a] p-values in this table are 2-tailed since there is no *a priori* hypothesis about which group has higher means.

Table 3.3

MZ and DZ differences: age at first productive combination, rate of using combinations, and rate of using productive combinations

	MZ	DZ	t (p-value)[b]
age difference at 1st productive combo (in days)	22.7	48.7	-1.83 (.038)
co-twin difference in rate of new combos per day	.23	.69	2.07 (.023)
co-twin difference in rate of productive new combos per day	.12	.43	2.14 (.020)

Table 3.4

MZ and DZ correlations, heritability, and shared environment estimates for age at first productive combination, rate of new combinations, and rate of new productive combinations[19]

	MZ r	DZ r	h^2	c^2
co-twin r for age at 1st productive combo	.96	.33	.90 t = 6.92, p < .001	0
co-twin r for rate of new combos per day	.93	.17	.84 t = 5.25, p < .001	0
co-twin correlations of rate of new productive combinations per day	.95	.03	.81 t = 5.06, p < .001	0

[b] p-values in this table are 1-tailed since we predict that MZs will have smaller difference scores than DZs.
[19] c^2 was constrained to zero in all three rows of this table to fix h^2 to lie between 0 and 1.

3.4.3 Summary

For all three measures (age of onset, overall rate, and productive rate), word combinations appear to have considerable heritability. Although actual h^2 estimates should always be interpreted cautiously, it is clear that the heritability of the onset of word combinations is relatively large and the shared and non-shared environments, c^2 and e^2, are quite small.

The actual sizes of the MZ and DZ correlations are worthy of comment though. For an additive genetic trait, the MZ correlation should be no more than twice the DZ correlation, yielding h^2 of no more than 1. In these data, the MZ correlations were more than twice as large as the DZ correlations, yielding $h^2 > 1$ and $c^2 < 0$ which had to be constrained to lie between 0 and 1 for sensible number to be estimated. There are three possible reasons for this result.

The first is some kind of error in data collection. Perhaps parents of DZ twins consciously or unconsciously emphasized co-twin differences, while pa rents of MZ twins emphasized similarities. Such a contrast effect has been seen before in parent report studies of non-cognitive measures, such as temperament. However, the lengthy review of the present methodology in chapter 2 presents several arguments against this possibility. First, in the vocabulary study, which utilized the same method of data collection, both MZ and DZ correlations were extremely high (> .9); no contrast effect was evident there. Second, both the CDI checklist and videotapes of parents making journal entries in the vocabulary study confirmed their relative accuracy. The CDI data indicated that MZ and DZ twins were confused to an equal extent. And although only MZ twins were videotaped, the data from those tapes indicated that parents did not artificially increase co-twin similarity to a large extent in journal entries for MZ twins. There is no reason to think that the word combination data, collected in exactly the manner—with parent journals—should be any different.

The second is error due to small sample size. Given the small number of twins, it is possible that the 19 MZ twins are unusually similar and/or the 17 DZ twins are unusually dissimilar, and these irregularities will work themselves out when more subjects are added.

The third possibility is a genetic explanation. Genes for the onset of grammar may not be just additive, but also dominant or epistatic. When more subjects become available, modeling techniques (e.g., Neale & Maes, in prep) can be used to determine the probability that combinatorial syntax is controlled by non-additive as well as additive genetic factors.

3.5 Nuisance variables

We saw in section 3.3 that several non-language variables were significantly different between the MZ and DZ samples: DZs had higher birthweight, higher gestational age, and higher SES (as indexed by father's education). They also had a lower mean age at reaching their first productive phrase. It is important to check whether any of these variables can account for the heritability results. Since only the age data showed and MZ-DZ difference in group means, only, this variables will be checked for their effect on the milestone data (i.e., the age at which the first productive combinations were recorded). The goal is to find out whether any of these variables had any effect on age difference between twins and more importantly, whether zygosity has predictive power when these other variables are partialed out.

For the 36 pairs of twins for which all the following data were available, a multiple regression was performed using absolute age difference as the dependent measure and using zygosity, average age, average birthweight, absolute birthweight difference, gestational age, and SES as independent variables. The results are in Table 3.4. Of all the nuisance variables, only average birthweight and birthweight difference were significant predictors of age differences at the first productive phrase. Furthermore, zygosity continued to be a significant predictor, so the genetic effects we found still stand.

Table 3.4

Variable	**B**	**SE B**	**Beta**	**T**	**Sig T**
Zygosity	40.40	15.93	.46	2.54	.017

Average age	.04	.10	.07	.44	.663
Average birthweight	-37.35	13.53	-.85	2.76	.010
Birthweight difference	15.59	7.84	.30	1.99	.056
Gestational age	8.86	5.50	.51	1.61	.118
SES	12.03	7.88	.25	1.53	.138

3.6 Motor milestones

As with vocabulary data, one might expect overall motor development to be highly relevant to the production of multi-word combinations. After all, the longer the utterance, the more motor planning and action required for successful execution. To test the possibility that the onset of word combinations is predicted by motor development, age differences at rolling over, and crawling were tested for correlations with the age differences at the first productive word combination. The purpose of these tests was to find out whether the onset of productive word combinations is a skill that might be more closely related to motor development than to cognitive or linguistic development (Lenneberg 1967, Ullman *et al.* 1997). Neither of these early motor variables had a significant correlation: the correlation between age difference at rolling and age difference at first productive combination was .12, and the correlation between age difference at crawling and age difference at first productive combination was .24.

3.7 Discussion

Although the number of twins in this study is smaller than that for the vocabulary study in Chapter 2 (36 pairs as opposed to 75), there seems to be a clear trend: the onset of productive word combinations is heritable. Unlike in the vocabulary study, in this study, h^2 is clearly above zero, and examination of the difference scores in t-tests confirms that there are genetic effects. This finding is unlike any before in the behavior genetics literature, since it shows directly that the very beginning of syntax—the onset of word combinations—is heritable. Furthermore, although we cannot make direct comparisons with confidence, the heritability of first word combinations appears to be a great deal larger than that for pure vocabulary growth in the one-word stage or the first 100 words. Recall that the largest heritability we found with vocabulary was .15, while for word combinations we saw heritability estimates of .81 to .91.

If the trend of higher heritability for word combinations than for single words is verified with the addition of more subjects over the next few years, it will be very exciting. One explanation for such an increase is that heritability tends to increase over age. Since children always begin the use of word combinations after the single-word stage, the larger heritability of word combinations may reflect the fact that kids are older. Increasing heritability over age could indicate increasing effects of already active genes, or the addition of new genes coming on line over time. Or it could reflect a decrease in importance of non-genetic factors over time.

Another possibility is that learning single words is a different process from combining them. To use single words, aside from planning and executing movements of the articulatory apparatus, the child must be able to identify and individuate objects and events and refer to them. To combine words correctly, the child has to know everything necessary for producing single words, but she also has to know the grammatical rules of word combination in her language. The addition of syntax makes combining words a fundamentally different process from saying them in isolation.

A trend related to the increasing heritability is the decrease in shared environment from single words to word combinations. Mathematically, an increase in h^2 does not require a decrease in c^2. However, since the MZ correlations stayed roughly constant between the vocabulary and word combination data, increasing h^2 does force a decrease in c^2. Intuitively, both the increase in h^2 and the decrease in c^2 ring true. The words children use are rather directly tied to their environment. Word combinations, on the other hand, require the same maturational factors, but they cannot be driven by the environment for two reasons. One, there is no environmental explanation for why it takes so long for children to start using phrase when they are ubiquitous in the environment (this is the triggering problem of Borer & Wexler 1987). Two, children rarely use novel words but frequently create novel combinations. Unlike individual words, phrases are not imitated—at least not once the child begins using them productively.

Figure 3.1

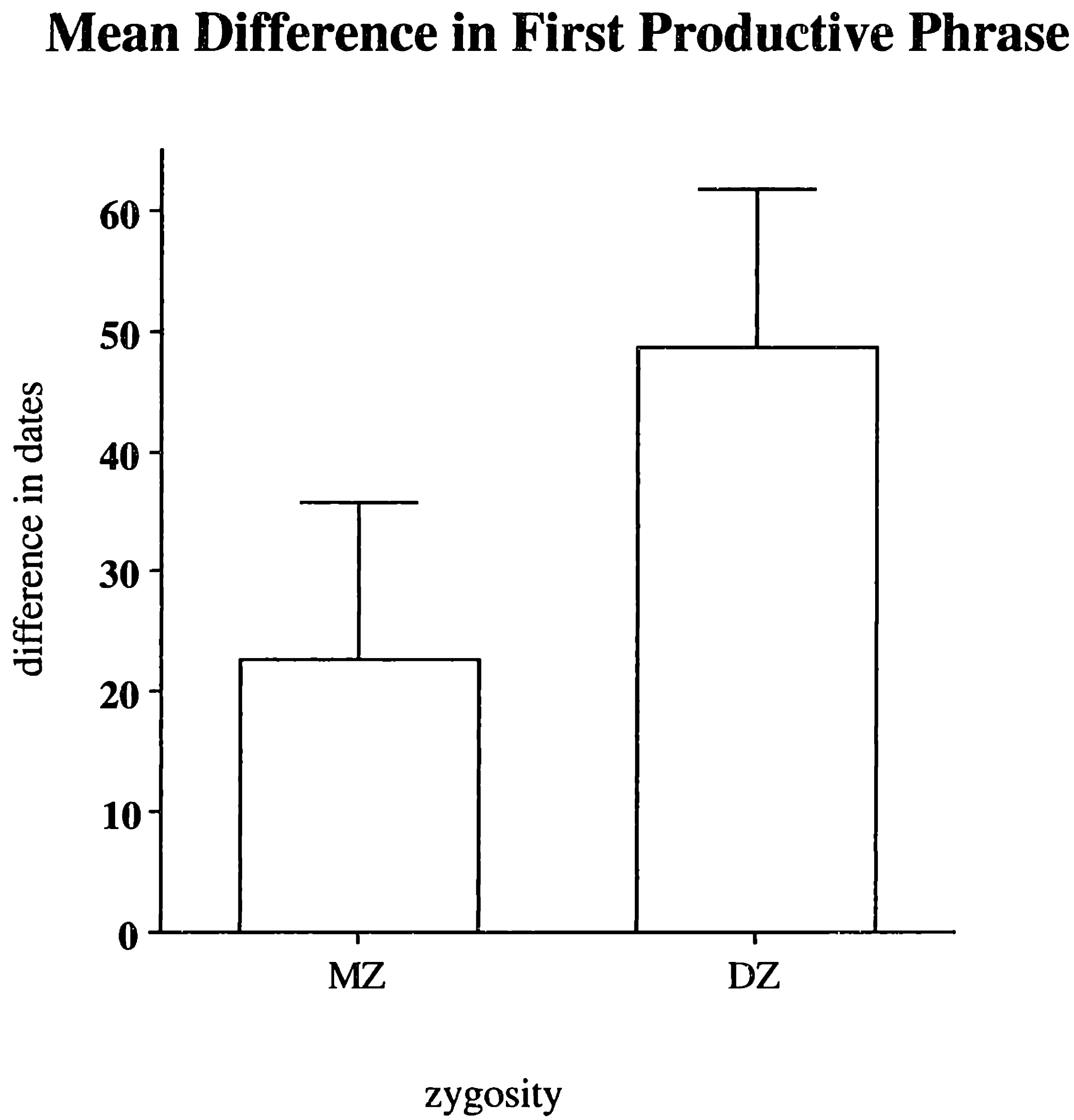
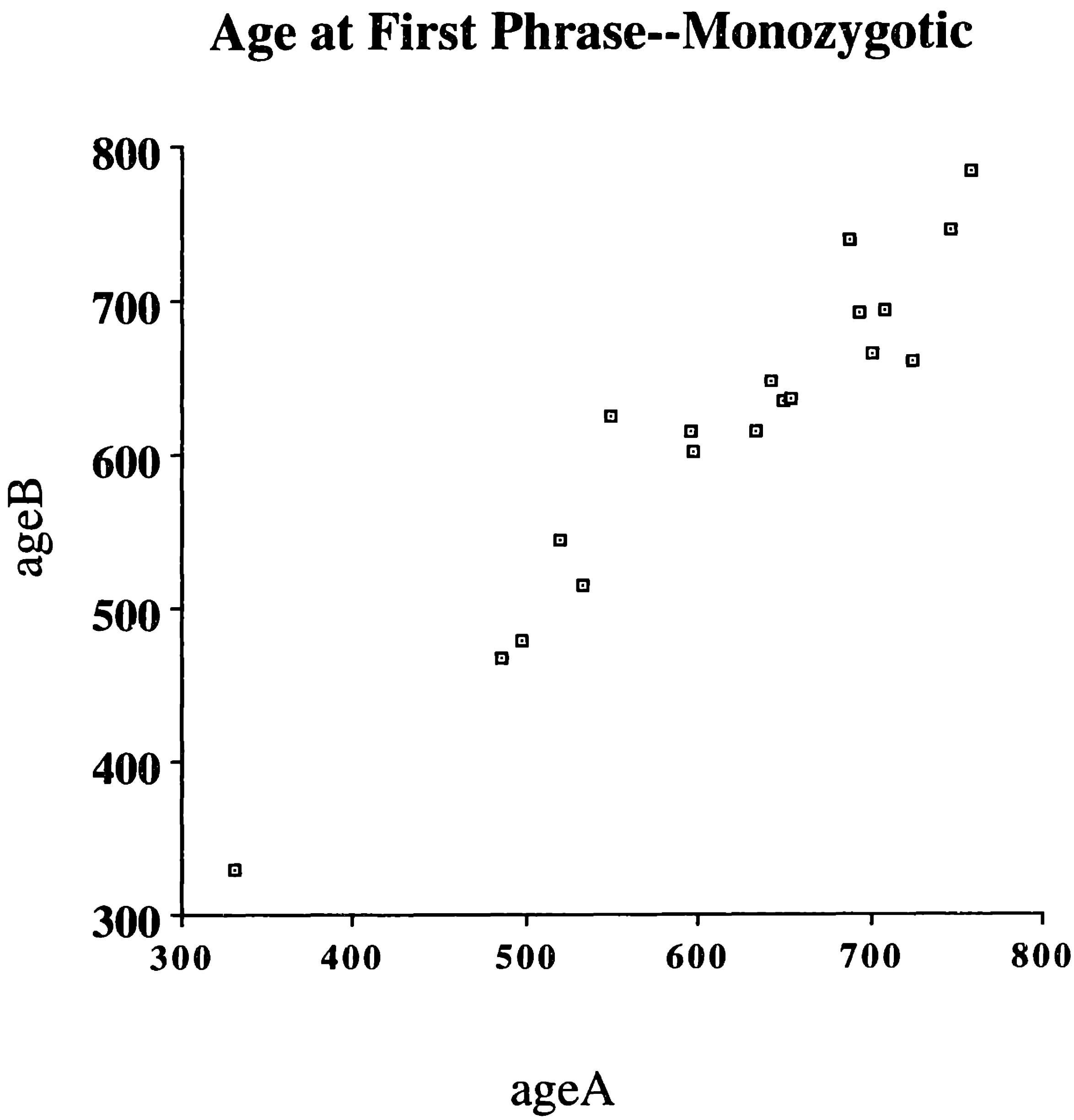

Mean Difference in First Productive Phrase
difference in dates
60
50
40
30
20
10
0
MZ
DZ
zygosity

Figure 3.2



Age at First Phrase--Monozygotic
800
700
600
500
400
300
ageB
300
400
500
600
700
800
ageA

Figure 3.3

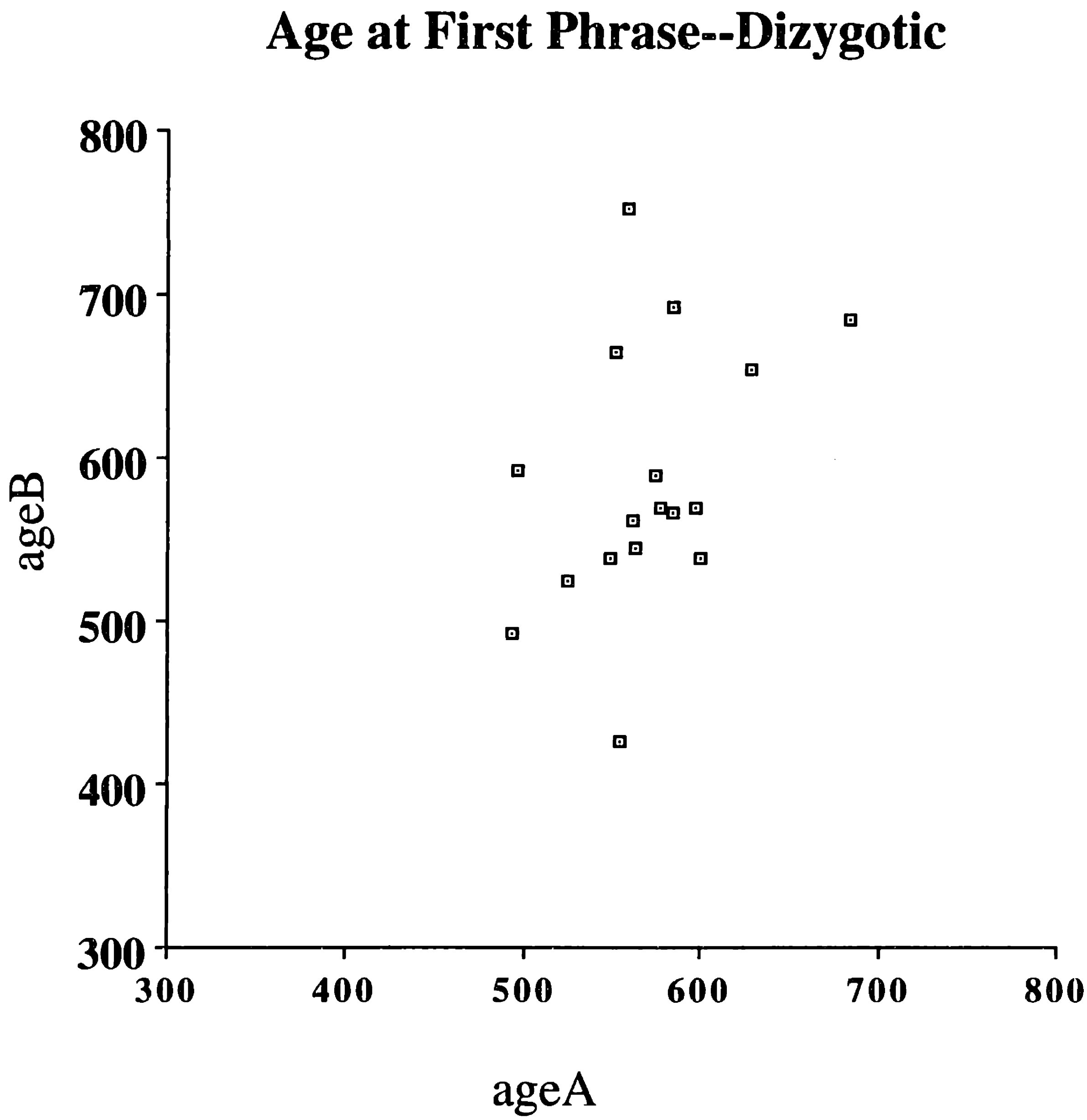

Age at First Phrase--Dizygotic
800
700
600
500
400
300
ageB
300
400
500
600
700
800
ageA

Figure 3.4

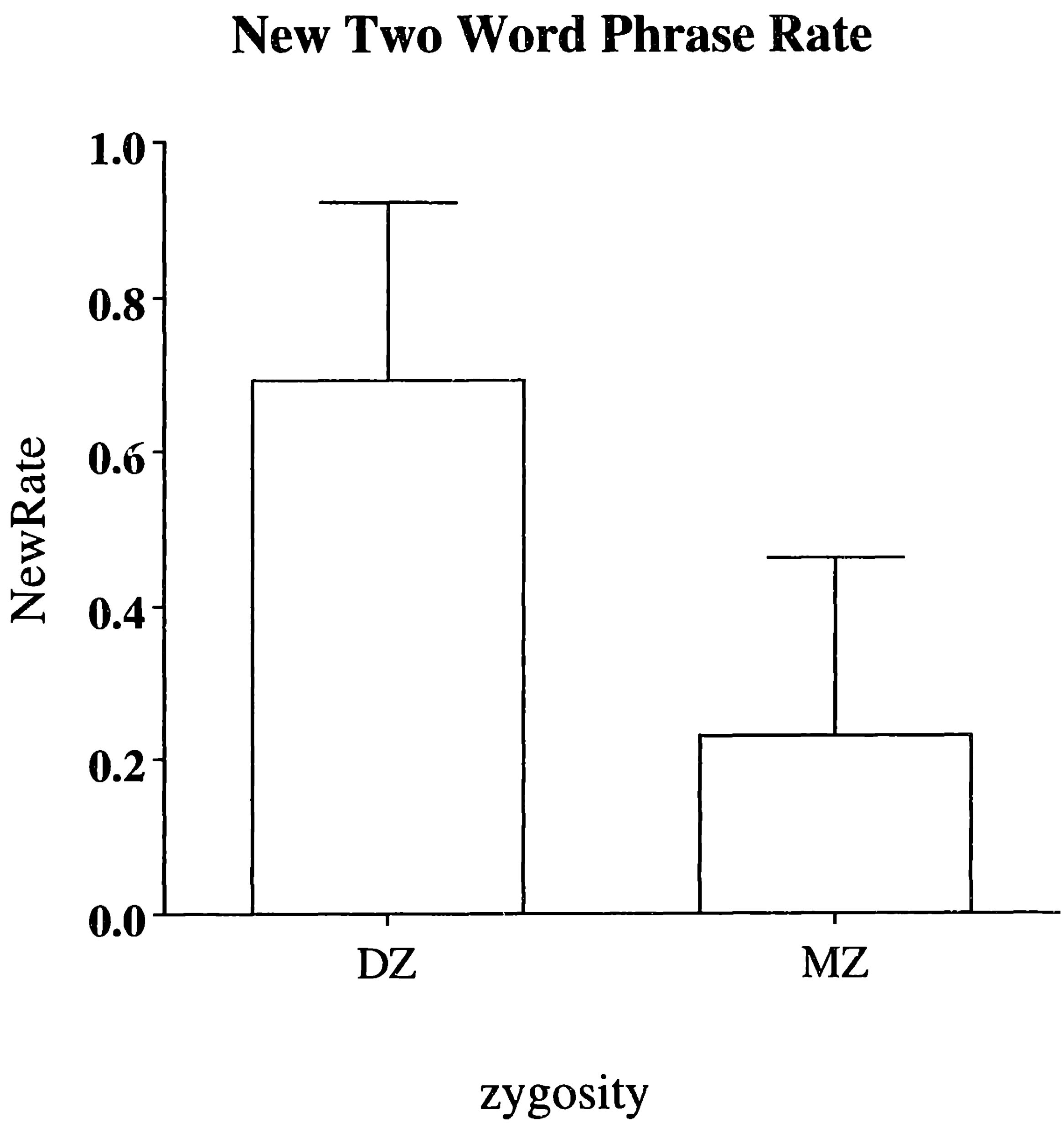
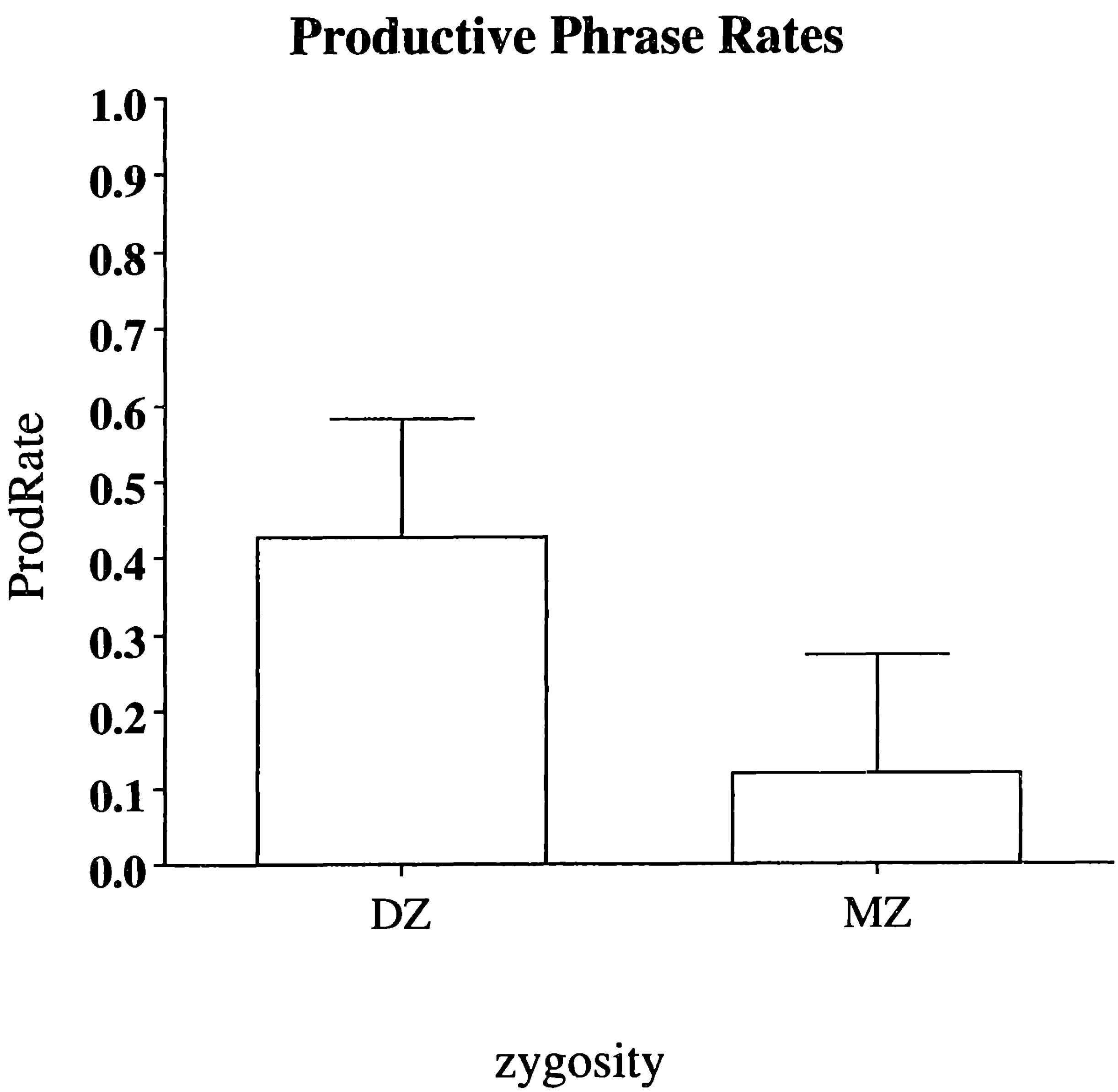

New Two Word Phrase Rate
1.0
0.8
0.6
0.4
0.2
0.0
NewRate
DZ
MZ
zygosity

Figure 3.5


Productive Phrase Rates
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
ProdRate
DZ
MZ
zygosity

Chapter 4

Past tense overregularization[20]

Early word combinations of the "more juice" variety are the child's first forays into syntax and as we saw in the last chapter, variation in the onset of these utterances may be driven to some extent by genetic variation. The next two chapters (this and the next) are devoted to more specific and adult-like examples of grammar. The first, to be covered in this chapter, is overregularization of the past tense. The second, which the next and final chapter is devoted to, is the development of verb tensing more generally. This chapter begins with a review of variation in syntactic development, then reviews behavior genetic studies of syntax, and then closes with the overregularization study.

4.1 Variation in syntactic development

Although the acquisition and final state of syntax may be idealized as highly uniform on a qualitative level, one can easily find quantitative variation in the timing of syntactic development. It is the goal of this work to look for genetic and environmental influence in this variation. One potential obstacle to this goal is that syntactic development takes place very quickly. Wexler (1996) has even suggested that many parameters are set before the child begins producing multi-word utterances, and that many more are set soon after (so-called Very Early Parameter Setting, or VEPS). Others have suggested that most syntactic learning is completed by the age of 4 years and what remains to be learned after 4 has more to do with social or logical aspects of language (i.e., the field of study known as pragmatics) than with syntactic competence. To attenuate this problem, the focus of this study will be on children between the ages of 2 and 4. As in other parts of language acquisition, there is ample evidence for quantitative individual variation in syntactic development both generally and in the specific topics of interest here.

[20]The twin data reported in this chapter were collected and analyzed in collaboration with Allison Baker.

Roger Brown's seminal study of three children acquiring English (Brown 1973) is *prima facie* evidence that there is individual variation in syntactic development. The book describes only three children, but they vary by as much as 16 months in acquiring grammatical morphemes. For example, one child, Eve, acquired regular past tense morphology at age 2;2, while Adam did not reach Brown and Cazden's criterion on past tense until 3;6 (see Brown 1973; Cazden 1968 for details). More general individual differences in syntactic development have been documented by Bates, Dale & Thal (1995), who note that children in the CDI norming sample varied in age by 6-12 months when equated for their mean length of utterance (MLU), as indexed by parental report of the use of grammatical morphemes by their children. The variance in age means that for a fixed MLU, which is an index of inflectional complexity and overall sentence length, there is a wide range of ages. Furthermore, nearly any report of syntactic development in children will show that they are not in perfect synchrony, although the development of certain constructions may consistently precede others in all children (e.g., Brown 1973; Snyder & Stromswold 1997).

4.2 Review of twin studies

Despite the fact that there is much variation in the syntactic development of children, and despite decades of debate over the relative importance of genetic and environmental influences in language learning (e.g., Snow & Ferguson 1977; Gallaway & Richards 1994), there are relatively few twin studies of syntax, and virtually none with children under the age of three[21]. The few studies that have been reported indicate that syntax is heritable, as measured by a variety of instruments. These studies are reviewed below and the reader is referred to Stromswold (under review) for a more thorough review and meta-analysis.

There are, to the author's knowledge, seven relevant studies of language in young twins, two of them unpublished. Many of these use standardized tests which conflate a variety of language abilities and peripherally related abilities (like verbal reasoning), and few test children

[21] I am considering here only twin studies conducted for the purpose of studying heritability. There are in fact numerous other twin studies which aim to study the special circumstances of being a twin, the "secret languages" of twins, and language delays in twins.

under 3 years old. As we saw in Chapter 2, the verbal component of IQ shows heritability, but verbal IQ measures a combination of tasks only peripherally relevant to grammar, such as logical reasoning, vocabulary (often assessed by asking for definitions), world knowledge, and verbal short term memory. We would like to make our measure of language as syntactically relevant as possible. Tests more specific to grammar might include comprehension tasks, in which comprehension depends crucially on the understanding of a particular morpheme or other syntactic device, or production tasks, where a context is clearly established and the presence versus absence of a particular morpheme required in that context is assessed (e.g., presence of past tense morpheme (*ed*)on a verb in a context that calls for the past tense of a regular verb). Attempts in this direction come from Munsinger & Douglass (1976), Mather & Black (1984), Mittler (1969), Fischer (1973), Chubrich (1971), Thompson, Detterman & Plomin (1991), and Reznick, Corley & Robinson (1997).

Munsinger & Douglass used two tests, the Assessment of Children's Language Comprehension (ACLC, Foster, Giddon & Stark 1972) and the Northwest Syntactic Screening Test (NSST, Lee 1971), with 8- to 10-year old twins. The ACLC consists of forced choice trials where the child chooses one picture (out of two or more similar pictures) that matches the sentence they hear. For example, in the syntax portion of the test, there might be a picture of a woman sleeping and a woman standing, and the experimenter asks the child to point to the woman who is sleeping. The NSST is similar, but has both comprehension and production tasks. The authors found significant heritability (.79) in their 74-twin pair sample, but it not clear what we should make of this since the twins in their sample are older than 4 and have therefore completed much of language development. It is likely that these children commanded all the grammatical constructions they were being tested with in their spontaneous speech, but could be forced to make errors when task requirements were increased. Success with these task demands may be more related to social competence, working memory, and overall IQ than grammar. Thus in terms of the heritability of syntactic development, it is not clear how to evaluate this result. It does, however, indicate the heritability of a heavily syntax-dependent task.

Thompson, Detterman & Plomin (1991) administered a verbal achievement test (a subtest of the Metropolitan Achievement Test, Prescott *et al.* 1986) and a verbal ability test (a test developed at the Colorado Adoption Project) to 278 pairs of twins age 6-12 years. Both language measures were found to be heritable; $h^2 = .54$ for the ability test and .27 for the achievement test. Although these results again indicate that some language related abilities are heritable, it is impossible to tell whether these tests are just another reflection of the fact that verbal IQ is heritable as opposed to a demonstration that syntactic learning is heritable. Furthermore, the age range of the twins is, once again, a little bit higher than what we are interested in here.

Mather & Black (1984) came closer to the age we are interested in (79 pairs of twins ages 3 to 6 years, mean age 4.5) and used two syntactic measures. The first was the Grammatical Closure subtest of the Illinois Test of Psycholinguistic Abilities (ITPA) (McCarthy & Kirk 1961), which "assesses the child's ability to produce grammatical inflections to complete statements about pictures" (Mather & Black: p. 304). The second was the Berko test (Berko 1958), which prompts children to give inflectional endings for nonsense nouns and verbs. Mather & Black reported that neither of these is heritable. However, these authors do not describe the procedure they used to test for heritability. Examination of their raw correlation coefficients, reproduced in Tables 4.1 and 4.2 below, indicate a heritability of .56 for the ITPA-GC and .10 for the Berko test ($c^2 = .12$ and .87, respectively; $e^2 = .32$ and .08). Without raw data, it is not possible to provide error terms for these estimates using the DF method. However, a Z-test comparing the MZ and DZ correlation coefficients for each measure revealed that the differences between MZ and DZ correlations are in fact significant.[22] Thus there is some heritability for understanding and producing grammatical inflections for real and nonsense words for children age 3-6.

[22] This test was made following the procedure outlined by Snedecor & Cochran 1989. The method is to convert the *r*'s to *Z*-scores (using the Fisher transformation or a published table) and divide the difference between them by the sum of the error terms (1 / (n - 3)) for each *Z*), where n in this case is number of twin pairs.

Tables 4.1 and 4.2

Tests of significance of the difference between MZ and DZ correlations for the ITPA-GC and Berko tests from data reported in Mather & Black (1984).

ITPA-GC	n (pairs)	r	r—>z	Z of difference	p
MZs	50	0.68	0.83	1.66	0.049
DZs	29	0.40	0.42		

BERKO	n (pairs)	r	r—>z	Z of difference	p
MZs	50	0.97	2.09	2.06	0.020
DZs	29	0.92	1.59		

Mittler (1969) looked at twins at about the same age, testing 200 four-year old twins using the entire ITPA. Mittler reports a heritability of .46 for the grammar subtest and heritabilities ranging from .36 to .56 for the other three subtests and the composite (similar to the heritability estimate for the ITPA we obtained by re-analyzing Mather & Black's data). Stromswold (under review) confirms that the MZ-DZ difference is indeed significant for the grammar subtest. So, once again, a test with young children that is somewhat specific to grammar shows heritability.

Fischer (1973) tested 21 pairs of twins (10 MZ, 11 DZ) at an age at which children are probably still actively acquiring the fundamentals of syntax, 2.5 to 3.5 years. The syntactic measures she used included two tests of sentence repetition which are scored based on how many false insertions, deletions, and substitutions (of grammatical importance) are made; a test of comprehension of different tenses, voices, cases, and other grammatical processes using forced choice picture-to-sentence matching; MLU from a spontaneous speech sample; the Berko test; and another test of inflection which uses real words as opposed to nonsense words. Fischer reports that several aspects of sentence repetition had significantly higher correlations for MZ than for DZ twins, including number of insertions, deletions, and overall correctness in tensing verbs. She also found that sentence comprehension and MLU are heritable. However, no other tests, including the two tests of inflectional morphology, were found to be heritable. While the finding of heritability in several aspects of grammar is encouraging, the lack of heritability in specific tests of verbal morphology is surprising.

In another study of verbal morphology, Chubrich (1973) studied six pairs of twins longitudinally from about 3.5 to 4.5 years of age. He administered the Berko test and a picture matching sentence comprehension test of grammar to these twins each month. Unfortunately, because of zygosity uncertainty, there were only 2 DZ and 4 MZ pairs by the end of the study, so formal tests of heritability were not possible. The fraternal twin pairs appear to have larger differences over a period of 9 months than the identical twins do on both tasks but these did not come out in Chubrich's analysis (i.e., they were not significant). There were two potential problems with this study, though: the small sample size and the repetition of the same tasks with the twins once a month. The small sample size simply introduces a lot of room for error, while the repetition could work to artificially increase the correlations between both types of twins, since they learn to expect the task each month. Thus, while it is again disappointing that Chubrich did not find heritability for morphology in a probe task for 3-4 year olds, there could be many reasons why this was the case.

Finally, we should include another discussion of Reznick, Corley & Robinson's (1997) findings, reviewed in Chapter 2. Reznick, Corley & Robinson themselves make no claims about whether their tests have anything to do with syntax; they are included here to be thorough. In fact, recall the caveats that were noted in chapter 2 about their tasks. In their comprehension construct, which putatively includes some syntactic comprehension, many of the items may be understood based largely on experience because of the use of a small and fixed set of phrases in the test. Some of them, like *put 1 block on the paper* may require syntactic analysis to carry out, but others, like *where's mama* could be answered based on recognizing an overall structure and a few key terms. As we saw in Chapter 2, this receptive construct showed a heritability of .28 at 14 months (with c^2 of .34), .13 at 20 months ($c^2 = .52$), and .18 at 24 months ($c^2 = .51$), so its heritability was relatively small. The finding of low heritability in comprehension, especially compared to the other studies considered here, could be because Reznick *et al.* tested children at a younger age. However, it is more likely that Reznick *et al.*'s measures are just not very relevant to syntax after all.

In sum, these studies demonstrate that a variety of abilities that are arguably related to grammar are heritable. However, in some cases, the measures are so broad that it is unclear whether syntax is differentially tapped as opposed to broader language and cognitive abilities. Furthermore, several of these studies test children after the age when most of their grammar has been acquired, or they test children at one point in time so that actual patterns of development cannot be studied (with the exceptions of Chubrich 1971 and Reznick *et al.* 1997). If we want to argue that syntactic development is heritable, we need to study an aspect of syntax that one can argue on independent grounds is not obviously linked to general language ability or to general cognitive development, and we need to study children at a very young age. The present study focuses on overregularization of past tense morphology as a case study of learning a grammatical rule. Furthermore, it uses spontaneous speech data, so results are less likely to tap into experiment-specific attention and memory factors.

4.3 Overregularization of verbal inflection

Overregularization is the extension of a regular grammatical pattern to irregular or exceptional cases. With the past tense of English verbs, overregularization means overapplication of the regular past tense marker *-ed*, resulting in forms such as *falled* for *fell* or *break* for *broke*. All children go through a period of overregularizing the regular past tense marker *-ed* to irregular verbs (and even adults do it occasionally) (Marcus *et al.* 1992). However, there are noticeable individual differences in the length and robustness of this period. Of the children studied by

Marcus and his colleagues, Adam (Brown 1973; Cazden 1968) never displays more than 5% overregularization in a monthly session (i.e., 5% or fewer of his tensed irregular verbs are overregularized) and most of his overregularizations (percentage-wise) occur between age 4 to 5. Eve, on the other hand, has several sessions with 10-20% overregularization between age 1;8 and 2;2. Sarah is quite similar to Adam, showing a few overregularizations around 2;10 but reaching as high as 15% between 4;6 and 5;0. Finally, Abe averages around 20% overregularizations from 2;6 to 5;0, though with substantial variance.

4.3.1 Overregularization and heritability

The reason a behavior genetic study of overregularization is interesting is that overregularization may serve as case study of a grammatical rule, as Marcus *et al.* have argued. When children produce overregularizations like *breaked*, they are creating novel forms that they could not have heard before. They cannot generate such forms without a rule *add -ed* in their repertoire. Thus, a child's first overregularization is the latest point at which he can be said to have learned this grammatical rule. Estimating the heritability of this point may be a test case for the heritability of grammatical rules generally. Since the actual form of the rule *add -ed* is arbitrary (specific to English), naturally experience will drive individual differences to some extent. In other words, if a child has to hear *verb + ed* used a minimum number of times before s/he can begin to hypothesize its meaning, children will differ based on the number of times they've been exposed to *-ed*. However, once this threshold is crossed, children's rule formation procedures, however acquired, kick in, and they may be governed by innate abilities.

It could be argued that first overregularizations do not necessarily reflect the onset of the rule *add -ed*, since this rule could be learned earlier and just not used until a retrieval failure occurs with an irregular past tense. Furthermore, children have been observed to use correct regular past tenses before they use their first overregularization, meaning they might already have the rule. However, Marcus *et al.* (1992) argued persuasively that the onset of overregularization correlated with an increase in correct regular past tense marking. Both signs that children know the rule *add* *ed* seem to come in together, supporting the proposal that we can look to overregularizations to track the onset of the rule.

The implications for a twin study are clear. If learning the rule *add -ed* (and hence overregularizing) is governed mainly by frequencies—by the number of times a child has heard the rule being applied unambiguously—and differences between children in overregularization onset and rate are mainly driven by differences in input, then both MZ and DZ co-twins should begin to overregularize at around the same time and about the same rate. If, on the other hand, innate grammatical rule formation modules simply mature at different rate in different children for genetic reasons, then MZ twins should mature at roughly the same time while DZ twins should be more discrepant on average. While this study has too few twins to be able to make any definitive conclusions, it is worth looking at the data so far.

4.3.2 Subjects

The twins in this study were recruited and screened in the same manner described in chapter 2. A total of 27 pairs of twins (16 MZ, 11 DZ) provided enough data to be used in the analyses below by the time of writing (i.e., they had used at least one overregularized verb). They participated for an average of 52.8 weeks and provided an average of 48 weeks worth of data (some of these subjects are still participating). The average age at the beginning of data collection for the overregularization study was 25.8 months and the average age at the end was 37.8 months, for a mean age of 31.8 months. Eight of these twin pairs (3 MZ, 5 DZ) previously participated in the vocabulary or word combination studies. The average birthweight of this sample was 5.85 pounds (5.88 for MZs, 5.81 for DZs). The average father's education fell at 3.1 for MZs and 3.2 for DZs (a rating of 3 means college was completed; see rating scale in Chapter 2). This information and other basic biological and sociological data are summarized in Table 4.3 below. The only significant difference between MZs and DZs is in co-twin birthweight difference, where DZs have a larger average difference than MZs, the same effect we have seen in each sample so far. Once again, the selection criteria for these studies probably eliminated MZ twins with large birthweight discrepancies, resulting in a sample that goes contrary to the typical twin population in terms of birthweight differences.

Table 4.3

Background data on the subjects in the overregularization study

Background variable	MZ	DZ	Difference
Female	12	6	
Male	4	5	
Total pairs	16	11	
Gestational age	37.3	37.7	t = -.43 (ns)
Birthweight (pounds)	5.9	6.09	t = -.39 (ns)
Birthweight difference (pounds)	.51	.97	t = -2.72 (p = .012, 2-tailed)
Weeks participated	55.4	40	t = 1.50 (p = .147, 2-tailed)
Weeks of data	50.3	37	t = 1.35 (p = .190, 2-tailed)
Father's education	3.3	3.2	t = .22 (n.s.)

4.3.3 Data collection and extraction

In order to track overregularization in twins, a very similar method to that used in the vocabulary and word combination studies involving parent report was used. For this study, parents were instructed to write down any sentence their twins used that referred to something that had already happened. They were told that we were interested in all ways of talking about the past, including correct forms like *talked* and *went*, incorrect forms like *goed*, and forms that are not marked at all, like *yesterday I go*. Instructions and report sheets are included in Appendix D. Parents are asked to keep these lists for an hour a day, as many days per week as possible, with 3 days per week being the minimum number acceptable. Initially, subjects were compensated with $50 or a zygosity test (cheek-cell swab DNA sample) after they'd been participating for four weeks. After the study had been in progress for about 1 year, the compensation was increased to $25 a month for new subjects in the hopes of retaining more subjects.

The data for this study were also kept on Microsoft Excel spreadsheets. The verbs from each child's sentences were categorized as to the kind of tense marking they had: regular stem form (unmarked for tense), irregular stem form, regular correct past tense, irregular correct past tense, irregular incorrect past tense (e.g., *brung*) and overregularized past tense (of several sub-types; see below). Repetitions and utterances with ambiguous contexts were excluded.

These data were used to calculate three dependent measures. The first measure was the age at the first overregularization. The second was the rate of overregularization, which starts for each child with his or her first overregularization and continues for the rest of the data available. The overregularization rate was calculated with the following formula:

Eq. 4.1

$$\frac{\text{\# overregularizations of all types}}{(\text{\# overregularizations of all types} + \text{\# correctly tensed irregular verbs})}$$

The different types of overregularizations are: irregular stems + ed (e.g., *breaked*, *runned*); irregular tensed forms + ed (e.g., *broked*, *ranned*); doubly marked irregular forms (e.g., *breakeded*, *brokeded*); no change verbs + ed (e.g., *putted*, *letted*); and novel verbs + ed (e.g., *spidered*, *lighteninged*). The third measures was tensing rates, ignoring overregularizations. These were also computed and compared within twins. Regular and irregular tensing rates were computed separately with the following formula, so that these tensing rates make reference only to obligatory contexts for tensing.

Eq. 3.2

$$\frac{\text{\# tensed forms}}{(\text{\# tensed forms} + \text{\# stem forms})}$$

Like the vocabulary study described in Chapter 2, it is difficult to know how accurately parents kept these records and in particular if they are zygosity-biased (see Chapter 2 for discussion), but we can look at three clues to test the general accuracy of parents. The first is simply to say that the results from the supplemental data concerning parent report of vocabulary reported in chapter 2, including videotapes, CDI records, and multiple raters support, in principle, the validity of the parent report past tense measure as well. The second is to note whether parents are reporting any stem (unmarked) verb forms. Intuitively, these are the hardest forms to convey the importance of to non-linguists. If parents are reporting stem forms, then it provides some reassurance that they also pay attention to tensed forms with some degree of accuracy. The third is to compare the ages of overregularizing, the rates of overregularization, and the rates of tensing that the parents report in this study to the same measures obtained by other methodologies. The second and third types of evidence were investigated below.

First, parents did report stem forms in abundance. The mean percentage of stem forms (relative to all other useful verbs provided by parents) was 28%. Therefore, the parents understood that they were supposed to write down stem forms, giving us some confidence that they performed adequately with the easier tasks (writing down tensed and overregularized forms). However, we suspect that the number of stem forms reported was slightly below what the children really used, because as we shall see below, the overall tensing rate of these twins (which is the complement of the stem rate), is high compared to non-twins. Parents understood the hardest point of the instructions—to judge when the context is past tense and to write down a stem form if the child uses it in this context. However, as untrained observers, they may not be noticing all the stem forms their children use, either because they are not salient enough or because it is not always possible to definitely identify a past tense context, especially if a stem form is used. Note that underreporting of these unmarked forms would not affect rates of overregularization, since, as the formula above shows, this rate is based only on overregularizations and correctly tensed irregulars. The decrease in unmarked forms only affects overall tensing rates, we will see shortly.

Turning to comparability of ages and rates of overregularization, the most complete estimates of overregularization ages (i.e., age at first overregularization) and rates are published in Marcus *et al.* (1992). These researchers investigated overregularization from the transcript data of 10 children from the CHILDES database (MacWhinney & Snow 1990) and also summarized the literature describing 15 other children and results from three additional studies (Gleason 1980; Gathercole 1979; Warren-Leubecker 1982).

The age at the first overregularization in Marcus *et al.* is reported for 7 children, garnered from their transcripts. The first overregularization occurred at a mean age of 2.4 years (median 2.5) for these children. For the 25 pairs of twins eligible for the age at first overregularization analysis in the present study, the mean age was 2.8 years (median 2.4). The twin average is just 4 months behind the Marcus *et al.* subjects, and the twin median is in fact .1 years ahead of them. Thus, parents are providing us with comparable estimates.

Marcus *et al.* report that the mean rate of overregularization was .073 (median .058) for the transcript children only, and for all the children (representing a mix of methodologies), the mean rate was .042 (median .026). In the twin sample in this study, the mean rate of overregularization (n = 24 pairs) was .215 (median .110). While this rate is higher than the mean from the Marcus *et al.* children, it is comparable to one subject, the child Abe. Interestingly, Abe has in common with our subjects that his parent (S. Kuczaj) kept his records. It would, therefore, be reasonable to conclude that the parents in the current study are providing overregularization data comparable to that in previous studies but that parent report provides a higher estimate. This may be because parent report is more accurate, perhaps because parents have access to a wider range of contexts, or because they have more information to judge past tense contexts (as opposed to an experimenter reading a transcript). On the other hand, parent report may be biased on the side of noticing overregularizations (which are arguably salient). All of these possibilities would lead to higher numbers of overregularizations[23]. However, there is one more possibility: twins may

---

[23]Marcus, p.c., confirms the trend of parent report data providing more overregularizations with observation of new parent-journal data he has collected

overregularize more because they have each other as models, a problem we will call the "twin effect."

Ideally, it would be possible to address the twin effect by checking for birth order effects on rates of overregularization in the existing literature, looking for a positive relationship between birth order and magnitude of overregularization rate. If hearing an older sibling overregularize the past tense of verbs causes a child to overregularize more, then the twin effect is something to be concerned about. Unfortunately, the children whose overregularizations have been studied in the most detail (Adam, Eve, Sarah, and Abe) appear to be only children (i.e., they had no siblings) for the period in which they were studied (Brown 1973; Kuzcaj 1977).

Another way to check for the twin effect is to remove from consideration those overregularizations that have already overregularized by a co-twin. Therefore, for each twin, the rate of overregularization was re-calculated for only those overregularized verbs that had not been used on a previous day (or the same day) by his or her co-twin. This rate consisted of the number of overregularizations a twin used that his co-twin had not used previously, divided by that plus the number of irregular correct verbs uttered after the point when the first non-overlapped overregularization was used, using the reasoning that an overregularization rate should not begin until the first "true" (i.e., non-overlapping) overregularization is used. The average recalculated overregularization rate was considerably lower than the original, at .120 (median .098), versus the original rate of .215, but some of this reduction is to be expected since removing overlapping overregularizations had a greater effect on the numerator than the denominator. The non-overlap mean was still higher than the Marcus *et al.* mean rates of .073 and .042, but less so. Further discussion of the twin effect and its implications are postponed for the moment, but the overregularization rates provided by parents seem to be reasonably comparable to standards in the literature, with or without removal of overlapping overregularizations, if we take into account that they were collected via parent report.

Finally, overall tensing rates in this sample of twins was also compared to rates reported in the literature. Rice & Wexler (1996) provide norms for 3-year-old normal children for the rate of use of *-ed* (regular tense marking) as part of their study on the tensing impairment of Specific Language Impaired children, so this is the first point of comparison. Forty non-SLI children were tested. The mean rate of correct use of *-ed* in the past tense in spontaneous speech, Rice & Wexler found, was .48. Marcus *et al.* also provided tensing rates. They broke down regular and irregular tensing rates for each month of data for the children Adam, Eve, Sarah, and Abe. For the same age period as the twins in this study (approximately 26 to 38 months), these four children had an average correct regular past tense marking rate of .61, and an irregular marking rate of .74.

The twins from the present study (26 pairs), whose average age was 31.8 months, had a mean tensing rate for regular verbs of .73, which is fairly close to the regular and irregular rates reported in Marcus *et al.*, but much higher than the regular tensing rate published in Rice & Wexler. One reason for the difference between the Marcus *et al.* estimates and the Rice & Wexler norms might be sampling differences; Eve and Abe have been shown to be precocious children linguistically. Perhaps these four children are not representative of the population. Our twins were also recruited through rather select means, mainly Mothers of Twins Clubs and the internet. Rice & Wexler's normal controls, on the other hand, were recruited through preschools, so they may represent a less privileged sample. Another reason could be that the estimates in Marcus *et al.* were based on a large number of transcripts (though from only 4 children) while the Rice & Wexler estimates were based on a maximum of 4 sessions of 15-30 minutes in length. Another possible reason is that past tense contexts are not always possible to identify in written transcripts (or even by parents in our twin study, apparently), while the Rice & Wexler study used an experimental procedure that makes the context unambiguous. Whatever the reason for the difference, it is clear that the twins in this sample are on par with the children in Marcus *et al.*, but ahead of the children in Rice & Wexler.

4.4 Results

Recall that twins were compared on three measures of the past tense: the age at the first reported overregularization, the overall rate of overregularizing, and the rate of correct tensing without overregularizations (both regular and irregular verbs).

4.4.1 Age at first overregularization

The age of each child was noted at his or her first overregularization of any type. The differences between co-twins' ages were used for a *t*-test, and the actual ages were entered into correlations. The average age of producing the first overregularization for MZs was 28.2 months, with standard deviation of 18.5 months; the average age for DZs was 28.9 months, s.d. = 22.6 months. Note the enormous variance of these numbers. The difference between co-twin ages was similar for MZs and DZs: 33.6 days (1.2 months) for MZs and 32.9 days (1.1 months) for DZs. The variation was very large here as well: standard deviation was 9.7 months for MZs and 7.1 months for DZs. The means were not significantly different by t-test, nor did they show a trend in either direction by non-parametric test (Mann-Whitney U).

The DZ correlation was actually higher than the MZ correlation for age at first overregularization: r = .91 for DZs, .82 for MZs. This trend obviously yields a negative heritability, which is not possible. h^2 was therefore constrained to zero and a DF regression was performed to estimate c^2, which came out to .87 (standard error = .39, $t(24) = 2.26$, $p = .05$). e^2 was equal to .13. Given the large variance in this measure and a higher DZ than MZ correlation, it is hard to interpret the results. However, it does appear that shared environment is very important in controlling individual differences in the age at first overregularization. One possible explanation is the twin imitation effect, which might induce apparent overregularizations to happen sooner than they otherwise would in some children. It would also explain the high MZ and DZ correlations, and the small sample size might be responsible for the MZ correlation being greater than the DZ correlation.

4.4.2 Overregularization rate

The overregularization rate over the entire period under study (as defined in equation 3.2), may provide more insight into this process. The actual timing of the very first overregularization is affected by the sampling practices of the parents, by their quality of reporting in general, by the twin effect, and by the availability of past tense contexts where a particular irregular verb (one that the child does not remember the correct past tense for) is appropriate. Since the whole study takes place over a relatively short period of time (about one year for most children), the overregularization rate, starting at the first overregularization, might provide a more stable approximation of the onset of overregularization. While random errors may bias the estimate of the very first overregularization , they should even out over the course of a year. Examining the rate allows us to test the strength of the rule "add *-ed*" over time.

Three pairs of twins (1 MZ, 2 DZ) who were in the previous analysis (of age at first overregularization) had to be excluded from this analysis because they did not begin overregularizing until the very end of the time they were in the study and so their rates would have been based on a very short amount of time (just a few weeks) compared to the rest of the sample. For the 15 MZs and 9 DZs remaining, the average MZ rate of overregularizing was .21 (overregularized forms per day), and the average DZ rate was almost the same, at .22. However, the MZ twins had significantly smaller differences in co-twin rates: .04 for MZs, .09 for DZs (t = 2.13, p = .023 1-tailed). The MZ correlation was .97 and the DZ correlation .75. Using the DeFries-Fulker method, h^2 for overregularization rate was estimated at .44, with a standard error of .41 ($t(20) = 1.07, p < .15$). c^2 was estimated at .53 (SE = .36; $t = 1.47, p < .10$), and e^2 was .03.

4.4.3 Tensing rates

The rate of past tense overregularizations is a particularly good means of assessing the use of the rule "add *-ed*," since overregularizations represent novel applications of the rule. The use of regular past tense forms also indicates the application of the past tense rule though; it is just not as stringent a test since a child could produce these forms either from memory or by application of the rule. We have also checked for genetic and environmental effects on individual differences in tensing rates. However, since we know that tensing rates are inflated for the twins in this study, probably because of parents underreporting stem forms, these data should be taken lightly.

Sixteen MZ and 10 DZ twin pairs were eligible for this analysis. Regular tensing rates were calculated for the entire period of the data, not just from the first overregularization. Overregularizations were not included in this analysis. The mean MZ rate of regular tensing was .77, the mean DZ rate .68 (not significantly different, $t = .89$). The mean MZ difference in rates was .03, the mean DZ difference .06. Although these differences are slight and not significant by t-test ($t = 1.33$, $p = .097$), the difference between them is in the predicted direction. The MZ correlation is .97, the DZ correlation .95. A DF regression revealed $h^2 = .05$ (standard error = .23, $t(22) = .22$), $c^2 = .93$ (SE = .19, $t = 4.89$, $p < .001$), and $e^2 = .02$. If we take this results at face value, it indicates that, unlike the overregularization rate, shared environment plays a large role in the regular tensing rate and heritability plays almost no role.

Tensing rates were also computed for irregular verbs. Although these forms must be retrieved from memory, they will not be triggered unless an abstract past tense feature is applied, so the formation of irregulars also reflects a grammatical process. Unlike the regular verbs, though, the irregular tensing rate showed no MZ-DZ difference. The overall rate of tensing irregular verbs was .63 for MZs, .62 for DZs and the mean difference score for MZs was .05, for DZs .06 ($t = .23$). The MZ and DZ correlations were both r = .95. h^2 was equal to .02 by DF analysis (ns). c^2 was equal to .94, (SE = .22, $t = 4.27$, $p < .001$) and $e^2 = .04$. Perhaps not surprisingly, c^2 was again the most prominent factor. Since irregular past tenses must be learned from experience, they should act like vocabulary items, with, as we found in Chapter 2, low heritability.

4.4.4 The Twin Effect

Although the finding of some heritability in the overregularization rate is intriguing, especially compared to the finding of zero heritability in regular and irregular tensing rates, we must consider the possibility that these data were tainted by the twin effect. The twin effect, which is the influence of a twin on his co-twin's overregularization rate due to imitation, may be driving up overregularization rates in this study. Besides giving twins a slightly higher rate of overregularization compared to other children, this effect causes another problem: some of the twins' overregularizations may not really be overregularizations. If a twin hears overregularized examples like *breaked* and *comed* in his co-twin's speech because his twin has started to overregularize, how is this child to differentiate his twin's overregularized forms from correct regular forms like *walked* and *talked*? The potential problem is that the twin who hears his co-twin overregularize could produce *breaked* and *comed* without their being overregularizations from the child's point of view. These "false" overregularizations could make overregularization rates invalid, since they would be a mix of overregularizations and, from the child's point of view, correct regular past tenses.

To check whether the heritability result of overregularization rates still stands despite the twin effect, the analysis was re-done by calculating the overregularization rates for non-overlapping verbs only (as described in section 4.3.3 above). Interestingly, the average MZ twin-pair percent of overlap was significantly higher by t-test than the DZ overlap (53% for MZs with variance 7%; 34% for DZs with variance 7%). Recall that MZ twins were slightly more likely than DZs to use the same words in the vocabulary study of Chapter 2; the overlap in verbs might reflect the same effect.

The new rate for MZ twins was .09; for DZs it was .17. With the re-figured rates, co-twin difference scores were once again smaller for MZ than for DZ twins (.04 versus .07), but not significantly so ($t = 1.04$, $p = .155$, 1-tailed). Furthermore, the MZ Pearson product-moment correlation was greatly affected by the recalculated rates, going from .97 for overall overregularization rate to only .42 for non-overlapping overregularizations. The change was probably due to a reduction in variance in MZ rates, which went from .04 to .01. Without substantial spread, it is impossible to have a high Pearson correlation. The DZ correlation, on the other hand, increased from .75 to .84, producing a heritability of zero. Since the MZ variance was so small, two non-parametric tests of association, Spearman's r and Kendall's Tau (both based only on ranks; Siegel & Castellan 1988), were also computed for the MZ and DZ data. In both cases, the MZ correlation was higher than the DZ (Table 4.4). The difference is not overwhelming but it is at least coherent (i.e., heritability is positive). Of course, with so few subjects it is dangerous to make any conclusions at this time.

Table 4.4

Pearson and non-parametric correlations for MZ and DZ overregularization rates after co-twin overlaps are excluded

	MZ	DZ
Pearson r	.42	.84
Spearman r	.74	.71
Kendall's Tau	.60	.51

Although it is important to do something to control for the twin effect, removing all overlapping overregularizations may be too harsh, since it removes all the most frequent verbs children overregularize (*go*, *break*, *fall*, etc.). Thus the recalculated rate may be too low. To address the twin effect properly, more work is going to have to be done, including finding a way to compare the overregularization rates of specific verbs before and after they are overregularized by a co-twin to typical rates of overregularization of those verbs in the literature. Finding out the ultimate effects of having a twin on overregularizing will have to wait for more twins and much more work with the data.

4.5 Discussion

We presented several findings in this chapter. First, we found that the age at first overregularization had no heritability and very large shared environment. However, the extreme variance in that measure, and the detrimental effect that the "twin effect" might have on its interpretation, cast doubt on the validity of that finding. Second, we found that the overregularization rate had both significant heritability and shared environment components. Third, we found zero heritability and large shared environment for regular and irregular tensing rate, though we noted that these might be depressed by the parent report technique

The finding of both heritability and shared environment in the overregularization rate should be viewed with caution, primarily because of the small sample size, but also because of the "twin effect," which may inflate shared environment or, if it is worse for MZs than DZs, inflate heritability. However, should this result turn out to hold, the implications would be intriguing. The overregularization of the past tense rule *add -ed* may represent a model case of a grammatical rule. Heritability of the overregularization rate over a short period of time in such young children would tell us that variation in the onset and strength of that rule has a genetic component. If this trend holds, this finding would be the first demonstration of heritability in a specific grammatical rule, as opposed to heritability of verbal ability generally, and would suggest that genes have a role in the development of grammar. This finding is especially interesting in light of the lack of heritability of regular and irregular tensing rates. It could be argued that all irregular forms and maybe even some of the regular past tense forms are stored as lexical items. Modulo the twin effect, overregularizations cannot be stored in the lexicon. The contrast in heritability between regular/irregular tensing and overregularization could reflect a distinction in the lexicon versus syntax.

Another interpretation may exist for the set of results obtained here, though. Overregularizations occur not only because a child knows the rule *add -ed*, but also because his lexical memory fails: he must fail to retrieve the correct past tense before he can apply the regular rule. Since memory failure is a prerequisite for using overregularizations, we might expect the heritability of overregularization to be similar to that of vocabulary that we found in Chapter 2. Indeed, for age of first overregularization, and for overregularizations after co-twin overlaps were removed, the heritability was quite small and shared environment quite large, just as we found in the vocabulary study. Even for the unmodified overregularization rate, the shared environment was still fairly high. Perhaps overregularization should be considered not as a test case of a grammatical rule, but as another way to study vocabulary: through retrieval failure.

Chapter 5

The development of tense and agreement

In Chapter 4, the heritability of a particular grammatical rule, "add *-ed*," was tested. In this chapter the scope of syntactic development is extended to all verbal tense and agreement morphology in English.

5.1 Optional Infinitives

Traditionally, the development of verbal inflection has been viewed as a process of learning the correct bits of morphology in a paradigm (e.g., Brown 1973; Pinker 1984). For example, in English, a child must learn that *-s* is the third person singular present ending and the ending for all other person-number entries in the present tense paradigm is *-Ø* (i.e., no ending). Over time, children use the correct endings in their required contexts more and more, until they have apparently mastered them. Learning morphemes, on this view, is like any other instance of learning or association.

Wexler (1994) pointed out that there are at least three problems with this conception of learning inflection. First, the view of what children have to learn under the learning theory—which bits of inflection should be associated with which forms—is not in accordance with ideas in linguistic theory of what verbal morphology entails. Linguists have shown that there is a deep relationship between the overt morphology and the underlying syntax, namely, head movement of V to I (and sometimes to C) (e.g., Chomsky 1991, 1995; Bobaljik 1996; Baker 1988; Pollock 1989). A theory that refers only to surface forms falls short of describing the process of verbal inflection. Second, if children simply have to the learn exact forms of inflectional morphemes, the process seems to take an inordinately long time for the task, considering that children hear these forms all the time and can even produce them on occasion. Third, children show a pattern of errors which is not explained by the learning theory. Specifically, although they frequently omit inflections, they rarely misuse them—that is, they rarely produce an inflectional ending in the wrong context. For example, English speaking children frequently produce "he go" but rarely say "I goes." If children's learning is a matter of association, one might expect both sorts of errors (misses and false hits) to be common.

As a new way of addressing the empirical facts and the theoretical problems associated with the learning theory, Wexler proposed the Optional Infinitives (OI) model. On this theory, children know the underlying syntax of verbal inflection (i.e., head movement), they know that it is inextricably linked to the overt morphology, and they know the forms of the morphological endings. Their problem lies in the fact that the whole apparatus is optional. But whether verbal inflection is used or not, the OI model holds that verbal inflection should come as a package for the child just as it does for the adult: verb movement and the morphology of tense and agreement should be inseparable. This prediction has proved true in every language examined, including German, French, English, and many others.

In German and other so-called "verb-second" or V2 languages, verbal inflection corresponds with movement of the inflected verb to the second position in the clause. If the verb is not inflected, it instead appears in the final position of the clause (presumed to be its underlying position). Children learning German obey this correspondence from the earliest observable stages. When they use an inflected verb, it appears in second position but when they (for mysterious reasons) fail to inflect the verb, it remains in final position, producing a clause ungrammatical for the adult but strictly conforming to the morphosyntactic constraints of the language (Poeppel & Wexler 1993). Under a learning theory this behavior is unexplained. If children's failure to produce verbal inflection is due to memory failure, verb movement (to clause-second position) should take place nonetheless. Under the OI model, on the other hand, this behavior is expected.

French provides another example. In negated sentences in French, the negative particle *pas* comes <u>after</u> an inflected or finite verb, but appears <u>before</u> a non-finite or infinitive verb. This distributional fact is another case of verb movement, as in the German example of movement from final to second position in the clause. (In French, the verb does not move all the way up to C but stops at I, and *pas* falls between I and V (Pollock 1989)). Like German children, though, French children also obey the correspondence between inflection and position. When French children produce inflected verbs, those verbs appear before *pas*, but when they produce a non-finite verb in a main clause (again for mysterious reasons), it appears after *pas* (Pierce 1992). Once again, the OI theory predicts empirical facts that are unexplained under a general learning theory.

Even in English, where both verb movement and inflectional morphology are relatively impoverished in comparison to other European languages, there are some empirical facts concerning the distribution of verbs that are also predicted by the OI model. As in French, non-finite verbs in English appear after negation, while tense and agreement (Tns and Agr) are realized as *do* before negation. The OI theory predicts that English speaking children will either produce non-finite verbs after negation with *do* before negation (e.g., *he doesn't want that*), or they will omit *do* altogether and still put the non-finite verb after negation (e.g., *he not want that*). Crucially, they should never produce *he not wants that* or *he want not that*. Transcript studies by Harris & Wexler (1996) confirmed this prediction.

To date, every non-null subject language studied shows the same pattern: where there is a correspondence between verb movement and inflection, children obey the correspondence but frequently fail to apply both. Other languages examined include Danish, Dutch, Faroese (Jonas 1995a, b); Norwegian (Wexler 1990, 1994); and Russian (Snyder & Bar-Shalom 1998).

In addition to the empirical facts reviewed above, the OI model makes a new prediction that is grounded in linguistic theory. In English, *Infl* features (tense and agreement) may be instantiated as inflectional morphemes (third person singular *-s* and past tense *-ed*), but they may also be realized in the auxiliary verb *do* and *be* when the main verb cannot move to *Infl* (I will not review the arguments here but see Chomsky 1957, 1991; Emonds 1976; Pollock 1989 for details). When tense and agreement features cannot both be checked, *do* and *be* cannot be inserted. Thus, at the same rate as*s* fails to appear in child language, so should *do* and *be* be omitted. One piece of evidence supporting this prediction came from Brown (1973), who found that the three children he studied mastered all of these items within a few months of each other. More recently, Rice, Wexler, & Hershberger (in press), in a control condition of their SLI study of verbal inflection, tested 40 three- and five-year old children at three 6-month intervals and found that the growth curves of all the tense/agreement markers could be modeled well with the same parameters. Thus the prediction that all *Infl* morphemes come in together seems to be true.

The OI theory, then, constitutes a leap in our understanding of the development of verbal inflection. Children do not so much have much trouble memorizing specific bits of morphophonology, as once thought, but rather fail to use tense and/or agreement sometimes because they cannot check both features. Furthermore, the OI theory unifies all of the realizations of *Infl*—in English, they are third person singular *-s*, regular past tense marked *-ed* and irregular past tense, auxiliary *do* and *be*, and copula *be*. On an associationist theory, they each posed a separate set of features to be learned. Of course, the forms of each morpheme must still be learned under any theory, but the fact that they all are apparently "mastered" at around the same time was not explained well until now.

5.3 OI and genetic influence

Two questions that remain, though, are why Tense and/or Agreement are omitted by children at all and why this omission is optional. Wexler (in press) has proposed that the omission of Tense and/or Agreement lies in a grammatical property of child language: it prevents the subject DP's D-feature from checking (in the spirit of Chomsky 1995) more than once. In the adult grammar, the D-feature on a subject DP checks both Tense and Agreement features; checking those features causes the subject DP to raise from VP. Wexler proposes that children cannot check both Tense and Agreement with one D-feature because they have a "Unique Checking Constraint" which prevents the D-feature from checking more than once (perhaps ultimately due to a misunderstanding of the interpretability of D; see Wexler in press). Since Agreement and Tense cannot both be checked, a sentence can converge for the child if one of them is absent from the representation. And, since most *Infl* morphemes in English can only be realized if both Tense and Agreement are present (Halle & Marantz 1993), these morphemes will be absent in child language (Schutze & Wexler 1996)

To account for the optionality of the process, Wexler (in press) proposes that children are faced with one of two violations: allow a non-convergent sentence in which one feature is not checked, or allow an interpretively odd but grammatically convergent sentence that is missing either Tense or Agreement. This choice arises because both inflected and uninflected forms may be included in numerations (Chomsky 1995). (A numeration is a list of lexical items from which a particular grammatical representation in constructed.) That is, the child, just like the adult, may start with a numeration that includes a finite verb or a non-finite one (or the presence versus the absence of any tense/agreement marker). If the numeration contains both Tense and Agreement, it cannot converge for the child (since only one of these features may be checked), but he will allow it anyway because it makes just this one violation. If the numeration contains only one of Tense and Agreement, it cannot converge for adults. For a child, though, it can converge because only one feature can be checked anyway. The result will be interpretively strange, causing another kind of violation (lacking grounding in time or subject verb agreement), but it will be grammatically convergent. Wexler proposes that the child can live with either one of these two violations (an unchecked feature or a sentence that lacks Tense or Agreement). Since the forms that go into the numeration are random, the child will sometimes end up with one violation and sometimes the other, resulting in what appears to be the optional use of Tense and Agreement on the surface.

Wexler (1996, in press) further argues that the Unique Checking Constraint, and the length of time that children show it, are best explained by genetically driven maturation. Learning the bits of inflection and the parameters determining word order, Wexler suggests, are easy learning problems that are mastered by children very early. The UCC, on the other hand, could very well be a maturational stage that children have to outgrow.

Further hints that the OI stage may be genetically based is provided by studies of Specific Language Impairment (SLI), a disorder which leaves most aspects of cognitive functioning unaffected but which impairs some parts of language, including aspects of grammar and pronunciation. Rice & Wexler (1996; Rice, Wexler & Cleave 1995) have characterized at least one set of SLI individuals as having no linguistic impairment other than a period of Optional Infinitives that is extended by as much as 2-3 years beyond that of unaffected individuals (the so-called Extended Optional Infinitive period, EOI). SLI and other language disorders are believed to be genetically transmitted (Neils & Aram 1986, Tallal, Ross & Curtiss 1989, Tomblin 1989 for SLI; Pennington 1990 for dyslexia; see Stromswold (under review) for a complete review), so it is possible that a mutation in a gene or set of genes that ultimately control verbal inflection is responsible for SLI of the EOI type. Studies are currently underway to determine whether the EOI type of SLI is genetically transmitted; preliminary reports suggest it is (Rice, Rice & Wexler, in press).

It is both the maturational hypothesis of normal development and the EOI period of SLI that inspires the study of optional infinitives in twins. The individual variation required by the twin method is amply met by the development of verbal inflection, as it is in most aspects of syntactic development. Aside from numerous smaller studies, Rice & Wexler have provided detailed data on the development of verbal inflection in a large sample of normal children as part of the control group for the SLI children they studied. Such development is more protracted than was generally believed by researchers, and it shows a great deal of individual variation. Rice & Wexler's tests take into account several aspects of tensing in English, including the third person singular present tense and agreement marker *-s*, the past tense marker *-ed*, copula *be*, and the auxiliaries *be* and *do*. The table below shows the means and surprisingly large standard deviations for 40 three year olds for some representative morphemes. (Rice, Wexler & Cleave 1995; Rice & Wexler 1996). So while children are well on their way to mastering verbal inflection by age three and are generally thought to have mastered it by age 4, there is apparently enormous variation between children in the attainment of full competence.

Table 5.1

Rates of tense usage in obligatory contexts by normal (non-SLI) 3-year olds, separated by morpheme, in spontaneous speech (from Rice & Wexler 1996)

tense marker	mean (%tensed)	standard deviation
3s	61%	34%
-ed	48%	40%
be (copula & auxiliary)	70%	24%

However, the fact that a deviant version of a trait (i.e., language impairment) has a genetic basis does not necessarily imply that variation in the typical state is driven by genetic variance. It is possible that a genetic mutation causes the disorder, but in the normal state there is no genetic variation and all individual differences are caused by environmental modulation. It is exactly this possibility that we want to test.

5.3 Previous twin studies of inflection

Recall that some of the studies reviewed in Chapter 4 sought to test the heritability of verbal inflection. Three different studies reviewed in that chapter (Fischer 1973, Chubrich 1971, and Mather & Black 1984) used the Berko test (Berko 1958), and one of them (Fischer 1973) also used the Mehrabian test of inflection (Mehrabian 1970). The Mehrabian uses real instead of nonsense words but is similar in spirit to the Berko test. All of these authors report no heritability for these tests, placing doubt on the heritability of verbal inflection. However, it is important to note that the Berko test (and the Mehrabian) include other inflections besides verbal tense and agreement. For instance, they include the plural noun marker *-s* and the progressive aspect marker *-ing*. Since neither of these are tense or agreement markers, it is possible that they are not heritable while tense and agreement marking are. In fact, Rice & Wexler (1996) report that in their SLI sample, the affected children show no impairment on plural marking or progressive marking; their impairment is selective to tense/agreement. Another possibility is that the developmental course of non-*Infl* morphemes is so much faster that they are essentially mastered well before *Infl* morphemes are. Support for this theory can be found in Brown 1973, who reports that progressive *-ing* and plural *-s* were mastered many sessions before the verbal inflectional morphemes. So, it is not unreasonable to think that the Berko test is not a pure test of the heritability of inflectional marking and that the lack of heritability finding could have been diluted by the inclusion of non-*Infl* items.

The only pure test of the heritability of the acquisition of an *Infl* morpheme we have seen so far was in Chapter 4 of this thesis, where we found hints that the development of the morpheme *-ed* was heritable from studying overregularizations of it. This finding suggests that the development of other *Infl* morphemes may be heritable as well. Although we did not see any heritability in regular and irregular tensing in Chapter 4, we suspected that that was because of parents' inability to catch all stem forms. In the study presented below, tape-recordings rather than real time records were used to capture the relevant forms, and trained experimenters rather than parents did the transcribing, so if individual differences in the development of *Infl* have a genetic basis, there is no obvious reason that it shouldn't be shown in this study.

5.4 A twin study of optional infinitives

The goal of this study is to track the development of tense/agreement markers over time to see if their emergence and developmental course is more similar for MZ than for DZ twins. The high level of detail needed, and the need to strictly control the context of children's utterances to separate marked from unmarked forms of verbs, meant that parent report was not an option for this study as it was for the other three studies in this thesis. In order to track the development of tense morphemes in detail over time, a very different approach was adopted. We followed 8 pairs of twins in the Boston area for 1-2 years, making audio recordings of their spontaneous speech one or two times a month. Although using only 8 pairs of twins prevents us from estimating heritability, it allows for the emergence of trends over time that would not be possible with just one or two sessions.

5.4.1 Subjects

The 8 pairs of twins, 4 DZ and 4 MZ, were recruited through local Mothers of Twins Clubs and local newspaper ads and were screened for the same factors as were described for subjects in Chapter 2. They lived in the greater Boston area within a 45 minute drive of Cambridge, Massachusetts.

Table 5.2

MZ twins

	Age at 1st session, in months	Age at last session	Number of sessions	Sex
MZ1	28.83	44.80	25	male
MZ2	24.77	41.60	30	male
MZ3	22.37	46.33	32	female
MZ4	29.40	56.07	43	female
Average	*26.34*	*47.20*	*32.50*	

Table 5.3

DZ twins

	Age at 1st session, in months	Age at last session	Number of sessions	Sex
DZ1	34	48	24	male
DZ2	28.23	45.53	22	male
DZ3	26.83	50.70	42	female
DZ4	32.57	42.47	16	female
Average	*30.47*	*46.61*	*26*	

5.4.2 Method

The twins were audio tape-recorded and occasionally videotaped in their own homes every two to four weeks for thirty minutes each. Each twin was recorded separately with his or her mother, another family member, or a second experimenter while the primary experimenter took notes on the session and operated the recording equipment. During the taping session, the primary experimenter also took notes on the context of the child's utterances, with particular attention to the timing of actions relative to the child's utterances. For instance, if a child said "it fall," the experimenter's notes would reflect whether that verb referred to an event that had already taken place (past tense context), an event that was taking place at the time of the sentence (present tense context), or an event that had not yet taken place (future context).

The tapes were then transcribed by the primary experimenter (the one who was present during the recording session), usually an undergraduate research assistant. The transcriber also coded the child's utterances using a coding system that reflected the time context, the presence or absence of inflection, the type of subject (e.g., pronoun, null subject, noun phrase, quantifier, etc.), the propositional type (declarative, imperative, or interrogative), and the type of verbal element. The verbal elements included auxiliary *do*, auxiliary*be*; lexical verb, non-finite verb, copula *be*, and modal. The coding was loosely based on the SALT system developed by Miller & Chapman (1991).

Several types of utterances were excluded from consideration in these data. Any child utterances which were direct repetitions or partial repetitions of another child or adult line were excluded. Phrases which a child repeated more than once in a transcript were excluded after the first utterance. Typical examples of these repeated phrases were *I don't know* and *what's that?* In addition, utterances in which the inflectional marking was ambiguous were excluded.

However, several utterance types that are sometimes excluded in other studies were included here. All contractions of *do* and *be* were counted as legitimate instances of those elements. Contractions are often excluded from consideration in child language studies because of the possibility that children memorize contracted words as unanalyzed wholes (e.g., *what's* and *it's*). However, we inspected the transcripts for evidence of this by looking for instances of words like *what's* and *it's* used in contexts where only *what* or *it* was required (e.g., *what's this?*), and found few. Since excluding contractions would have greatly reduced the number of examples available and since there seems to be no principled reason for doing so, they were counted. Other sentences that could have been excluded are those that contain simple present tense verbs in a context in which an adult would have used a progressive construction (e.g., *he go* or *he goes* when an adult would have used *he is going*). The justification for including these simple present sentences as such was that even though the child apparently got the semantics wrong, he still had an opportunity to tense the verb and reveal whether he had mastered the morphology. To check on whether this decision was justifiable, separate counts were completed for third person singular *-s* with and without these utterances and there was little difference between them, so there seemed to be no reason for excluding them.

Each transcript was proofread by the original transcriber, and in addition about 30% of the transcripts were proofread by a second person. The number and type of discrepancies between the original transcriber and the proofreader were noted. It turned out that in the 63 files that were proofread by a second person, there were a total of 4,118 lines of child speech that contained a code used in this analysis, and 277 or 7% of them contained a disagreement that would have affected the analysis. That is, there was a disagreement over the type of tense marker that was used or the presence or absence of it in 7% of the cases. This is a relatively low number, considering the subjectivity involved in transcribing child speech and especially in hearing inflectional endings and contracted forms of auxiliaries. Furthermore, since both twins in each pair were transcribed by the same person in almost all cases, these errors shouldn't affect co-twin comparisons. Of course, they will increase the similarity of twins in both types of twins, introducing rater bias, but as long as they are not greater for MZs than DZs, they will not artificially inflate the genetic effects we are looking for.

5.4.3 Measures

The coded transcripts were searched and organized using a program written by the author in the Perl language. The results consisted of a rate of correct use in obligatory contexts in each transcript for each of the tense/agreement makers. Rate of correct use is defined as the number of correct uses of an marker divided by the total number of contexts in which that marker should be used, or:

$$\frac{\#\text{times marker was used correctly}}{\#\text{times marker was used correctly} + \#\text{times it was not used when it should have been}}$$

The tense and agreement markers we studied were:

1 Third person singular *-s*, for example

He goe*s*.

2 Past tense *-ed*, as in

I walk*ed* to school.

3 Irregular past tense, as in

She *went* home.

4 Auxiliary *do* (both in questions and in declaratives with negation), as in

He *does*n't go, and

Does he go?

5 Auxiliary *be*, as in

He *is* going.

6. Copula be, as in

She *is* a student.

Separate counts were made for auxiliary *do* in questions and in declaratives, because it was discovered that the rates of correct use in questions and declaratives were often quite different even within a single child.[24] An average percent correct for all these elements was computed, weighting each element by the number of tokens it contributed. These percentages are provided for each twin pair in Appendix E.

Using the average percent correct as the dependent measure, each pair of twins was compared on changes in this measure over time using a series of correlations and regressions. For each pair, there were some 20-40 recording sessions over a period of 1-2 years. The average percent of correct tensing in obligatory contexts in each session for each twin served as the starting point for the correlations. However, some of these numbers, especially in early sessions, were based on rather small sample sizes and resulted in some wild fluctuations in tensing percentages. In an attempt to stabilize the data, instead of running correlations directly on the original numbers, we collapsed sessions covering a period of 1 month into a single number (based on the average of all the sessions in that month). This procedure resulted in numbers based on 1-3 sessions depending on how frequently the twins were taped (tapings were not always at even intervals due

[24]One possible reason for the difference is that in declaratives, *do* is only used when negation is present, either in the form of *not* or the clitic *n't*. The use of negation may force the presence of *do* because when the contracted form *n't* is chosen from the lexicon, it must be affixed to something in order to surface. The contracted form *n't* is more common in spoken language, so *do* will be forced in most cases where negation is present. Since negation is much more common in declaratives than questions (as a proportion), the use of *do* will appear to be much higher in declaratives. See Harris & Wexler (1996) for a similar suggestion as to why the correct use of *do* in obligatory contexts seems to be higher than the correct use of third person singular *-s* in obligatory contexts.

to scheduling constraints). Collapsing sessions reduced the total number of points that were entered into the correlation for each pair of twins, but it had the advantage of making each point more stable.

Thus for each pair of twins, there was a series of numbers representing percent of correct tensing at 1-month intervals. For example, 4 months' worth of data from one pair might look like the sample data presented in Table 5.4, except with 7 to 19 months instead of 4. Correlations were run between co-twins on these numbers.

Table 5.4

Example of tensing over time in a fictional pair of twins

Time interval	**Twin A**	**Twin B**
month 1	.50	.42
month 2	.67	.48
month 3	.66	.53
month 4	.70	.51

5.5 Results

We computed first order correlations between co-twins' tensing at each interval, resulting in a summary *r* for each twin pair. Then partial correlations were computed, controlling for the effect of age, which was linear (because the data were made to conform to one-month intervals). The partial correlations can tell us to what extent co-twins develop in the same way aside from the fact they are both monotonically increasing over time, an effect which drives up first order correlations. This analysis also provided a summary (partial) *r* for each twin pair. In order to compare the correlations (*r*) and partial correlations (*pr*) between MZ and DZ twins as groups, we first converted them into Z scores using the Fisher transform. All *r*'s, *pr*'s and Zs are given for MZ twins in Table 5.5 and DZ twins in Table 5.6.

Table 5.5

First order correlation coefficients (r) and partial correlation coefficients (pr) with the age partialed out and their Z equivalents for the MZ twins

	MZ r's	MZ Zs	MZ pr's	MZ partial Zs	n (months)
MZ1	0.84	1.22	0.57	0.65	10
MZ2	0.81	1.12	0.83	1.18	13
MZ3	0.78	1.04	0.01	0.01	15
MZ4	0.89	1.40	0.34	0.35	19
MZ averages		**1.120**		**0.55**	

Table 5.6

First order correlation coefficients (r) and partial correlation coefficients (pr) with the age partialed out and their Z equivalents for the DZ twins

	DZ r's	DZ Zs	DZ pr's	DZ partial Zs	n (months)
DZ1	0.81	1.12	0.21	0.21	12
DZ2	0.51	0.56	-0.64	-0.75	11
DZ3	0.63	1.04	0.08	0.00	18
DZ4	0.80	1.10	-0.21	-0.21	7
DZ averages		**.88**		**-0.17**	

As the tables show, MZs had higher first order correlations than DZs . The converted Z-scores were entered into a *t*-test and the difference between them reached statistical significance ($t = 1.99$, $p = .047$, 1 -tailed). The MZ-DZ difference in partial correlations was somewhat larger. These (as Z-scores) were also significantly different by t-test ($t = 2.20$, $p = .035$ 1-tailed) but with a bigger effect size. Both these findings are exciting considering there are only 4 pairs of twins in each group. The finding of higher partial correlations for MZs than DZs as a group indicates that whatever non-linearities (with respect to age) exist in the developmental curve of tensing over time have a genetic basis.

This is the first evidence for the genetic basis of variation in the development of the *Infl* system and as such supports the proposal that the optional infinitive stage is a maturationally determined one.

5.6 MLU[25]

It could be argued that the genetic influence revealed in the correlations between co-twins in tensing over time is due not to the maturing of the *Infl* system in particular, but to the genetic basis of the entire language system. If all aspects of language are more correlated in MZs than DZs, if overall grammatical and linguistic ability is increasing, then a result showing that the development of tense morphemes has genetic influence is not so exciting. As an attempt to show that the growth of tensing is not completely dependent on the overall growth of linguistic ability, we have computed the mean length of utterance (MLU) for each transcript. MLU is considered to be a very broad measure of language growth. The more a child's language develops, the longer his sentences get, both because the number of words is being increased and because the use of inflectional morphemes is becoming greater. We examined MLU for MZ-DZ differences and attempted to partial its effects out of the MZ-DZ differences in the *Infl* morphemes.

---

[25]I would like to express my gratitude to Melanie Soderstrom, an undergraduate research assistant, who single-handedly calculated these numbers. She also made decisions about which utterances were to be counted, coded them appropriately, and wrote Perl programs to search for and compile the codes. These calculations represent about 6 months of her work.

MLU was calculated for each transcript using lines that did NOT contain any of the following:

•a direct repetition of another utterance,

•a phrase that is repeated over and over throughout the session (e.g., *I don't know*),

•a one-word response to a question,

•any words or sentences that were interrupted by the child himself or by another speaker,

•any unintelligible speech or other sounds,

•elliptical utterances, or

•routines like a song or counting

All remaining lines were used. Certain fixed phrases, such as "thank you" and "boo-boo" were made into one morpheme and words that were repeated due to stuttering were counted only once. Each independent word stem and most inflectional morphemes (e.g., plural *-s*, possessive *-s*) were tallied as a morphemes. However, in order to make MLU maximally different from the tensing numbers we computed earlier, we excluded all tense morphemes (i.e., third person singular *-s*, *-ed*, auxiliary *do* or *be*, copula *be*) from the MLU counts. MLU was then computed by dividing the total number of morphemes from usable lines in the transcript by the number of usable lines.

Before attempting to partial the genetic effects of MLU out of tensing, it is worthwhile to check whether MLU itself shows genetic influences (recall that tense morphemes were not included in our MLU counts, so genetic influence of MLU would tap into a different underlying process than that for tense/agreement). Using first order correlations, MLU alone does not seem to have a more similar course of development in MZs than in DZs (a *t*-test performed on the Z-scores was not significant; $t = 1.32$, $p = .118$). As with tensing proportions, we also calculated partial correlations for co-twin MLUs with age (a linear component) partialed out. Unlike with tensing, though, the partial correlations were not significantly higher for MZs than DZs by *t*-test ($t = 1.14$, $p = .148$), though clearly there is a trend of MZs having higher *pr*'s than DZs, as should be apparent in Table 5.7. This trend suggests some genetic influence but further work will be needed (i.e., by adding more subjects or perhaps restricting the age range) before that can be determined for sure. These correlations are summarized in Table 5.7.

Table 5.7

MLU correlations and corresponding Z-scores and partial correlations with age partialed out and corresponding Z-scores

	MZ r	MZ r→ Z	MZ pr	MZ pr→Z
MZ1	0.80	1.10	0.11	0.11
MZ2	0.96	1.95	0.56	0.63
MZ3	0.55	0.62	-0.1	-0.10
MZ4	0.89	1.42	0.13	0.13
average		*1.27*		*0.19*
	DZ r	DZ r→ Z	DZ pr	DZ pr→Z
DZ1	0.78	1.05	0.61	0.71
DZ2	0.61	0.71	-0.47	-0.51
DZ3	0.82	1.16	-0.27	-0.28
DZ4	0.46	0.50	-0.67	-0.81
average		*0.85*		*-0.22*

In order to partial the effects of MLU out of the developmental course of optional infinitives, a regression was run for each pair of twins using the tensing data described above for Twin A of each pair (chosen by alphabetical order), averaged into bins of size 1-month, as the dependent measure. The independent variables were age, Twin B's tensing over time in the same 1-month bins, and Twin A's MLU (calculated without tense morphemes) also averaged over 1-month time periods. To summarize, the model equation was

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

where

Y = Twin A's tensing

X_1 = Twin B's tensing

X_2 = Twin A's MLU over time

X_3 = constant.

The idea of this regression was to see whether Twin B's tensing could account for any unique variance in Twin A's tensing. If Twin A's tensing is driven by the same factors that control MLU (something like overall linguistic growth), then A's MLU should predict tensing better than B's tensing. However, if tensing is controlled by an independent factor, and variation in tensing has a genetic basis, then co-twin tensing could predict as much or more variance in a child's tensing as his own MLU.

The results are shown in Table 5.8 (for MZ twins) and 5.9 (for DZ twins). In several cases, for both MZs and DZs, Twin A's MLU accounted for substantial variance in his own tensing over time. However, the important result here is that in many cases, Twin B's tensing still had quite a bit of predictive power on Twin A's tensing. Within the MZs, 3 out of 4 pairs had high coefficients for this variable; β_1 reached significance in all cases except MZ1, and in all cases accounted for a great deal more of the variance than Twin A's MLU. Within the DZ's, interestingly, results were more mixed. In only one case (DZ1) was TwinB's tensing a significant predictor of Twin A's tensing with MLU partialed out, and the other three cases MLU accounted for more of the variance than tensing.

Table 5.8

Results of regression, MZ twins

Dependent measure was Twin A's tensing

	Twin B's tensing (β_1)	Twin A's MLU (β_2)
MZ1	$\beta = .82$	$\beta = .02$
	$t = 1.94$, $p = .093$	$t = .04$ $p = .967$
MZ2	$\beta = .69$	$\beta = .41$
	$t = 4.78$, $p < .001$	$t = 2.88$, $p = .016$
MZ3	$\beta = .71$	$\beta = .10$
	$t = 2.63$, $p = .022$	$t = .37$, $p = .718$
MZ4	$\beta = .79$	$\beta = .12$
	$t = 3.14$, $p = .007$	$t = .43$, $p = .670$

Table 5.9

Results of regression, DZ twins

Dependent measure was Twin A's tensing

	Twin B's tensing (β_1)	Twin A's MLU (β_2)
DZ1	$\beta = .63$	$\beta = .31$
	$t = 2.94, p = .017$	$t = 1.46, p = .179$
DZ2	$\beta = -.32$	$\beta = .96$
	$t = -.64, p = .540$	$t = 1.92, p = .092$
DZ3	$\beta = .23$	$\beta = .46$
	$t = .58, p = .572$	$t = 1.14, p = .272$
DZ4	$\beta = .29$	$\beta = .727$
	$t = 1.43, p = .227$	$t = 3.5, p = .024$

It should be clear that for MZs, tensing is highly correlated between co-twins and MLU cannot account for this correlation. For DZ twins the picture is less clear; there appears to be a genetic effect on the extent to which a twin's tensing predicts his co-twin's tensing with MLU partialed out, since for DZs MLU has more predictive power for tensing than does co-twin tensing.

Thus, there was an MZ-DZ difference in the ability of tense alone, with MLU partialed out, to account for co-twin similarities. The implication of this finding is that individual differences in the development of *Infl* specifically, and not just grammatical or linguistic development generally, has a genetic basis.

5.7 Tensing errors: wrong agreement

Independently of the behavior genetic results above, it is worth noting another, non-genetic but important prediction of the Optional Infinitives theory: that children do not apply the wrong inflectional marking. Instead they either apply it correctly or fail to apply it. For example, in past tense contexts, children either fail to mark the past tense with *-ed* (on a regular verb) or mark it correctly. They do not apply *-ed* to verbs in a present tense context. Armed with this hypothesis, we looked for examples of misapplied inflectional marking in the twins' transcripts to find further evidence for the OI model.

Errors of the past tense ending *-ed* are actually rather hard to find because when a child uses *-ed* it is hard to prove beyond a doubt that he didn't have some past tense context in mind. Misapplications of the third person singular agreement marker *-s* are easier to look for because the contexts are clearer (a verb in a present tense context that comes after a subject that is not third person singular should not have an *-s* ending). Use of the wrong form of copula and auxiliary *be* are also easy to look for. Thererore we focused on these three types of wrong agreement.

We first counted the number of *-s* endings that occurred on present tense verbs with 1st person singular or plural, 2nd person singular or plural, or 3rd plural subjects and divided by the total number of *-s* endings on present tense verbs, calculating the following proportion:

$$\frac{\text{\# -s endings on 1st person, 2nd person, and 3rd person plural present tense verbs}}{\text{\# -s endings on 1st person, 2nd person, and 3rd person plural present tense verbs} + \text{\# -s endings on 3rd person singular present tense verbs}}$$

As predicted by the OI model, the number of misapplications of *-s* was very small (Table 5.10). The average proportion of *-s* that was misapplied was just .03. The highest proportion reached by any child was .11, and that is clearly because she has relatively few uses of *-s* in her speech.

Table 5.10

Proportion of applications of third person singular *-s* to other contexts

Family	child	misapplied -s	correctly applied	proportion wrong agreement
DZ1	A	5	273	0.02
	B	3	78	0.04
DZ2	A	4	130	0.03
	B	2	115	0.02
DZ3	A	1	29	0.03
	B	0	16	0.00
DZ4	A	1	141	0.01
	B	4	86	0.04
MZ1	A	0	98	0.00
	B	5	190	0.03
MZ2	A	3	369	0.01
	B	5	150	0.03
MZ3	A	0	68	0.00
	B	3	74	0.04
MZ4	A	1	28	0.03
	B	4	32	0.11

We next counted the proportion of incorrect forms of copula *be*. That is, given that copula *be* was used, what was the probability that the wrong form of the word was used? Note that it is common to use the wrong form in adult speech when a plural subject is used with a contracted form of *be*, such as *where's the cookies?* Examples like these were counted as incorrect in Table

5.11, so these numbers really represent an overestimate of what is truly ungrammatical. Remarkably, they are still quite small.

Table 5.11

Proportion of incorrect forms of copula *be*

Family	child	wrong forms	correct forms	proportion of wrong forms
DZ1	A	9	701	0.01
	B	3	374	0.01
DZ2	A	5	438	0.01
	B	2	538	0.00
DZ3	A	9	215	0.04
	B	1	83	0.01
DZ4	A	6	834	0.01
	B	1	645	0.00
MZ1	A	3	393	0.01
	B	3	429	0.01
MZ2	A	6	639	0.01
	B	11	589	0.02
MZ3	A	5	438	0.01
	B	2	538	0.00
MZ4	A	17	455	0.04
	B	18	405	0.04

Auxiliary *be* was examined in the same manner, calculating the proportion of incorrect forms given that a form of auxiliary *be* was used. The results are summarized in Table 5.12 below and again show that these errors are relatively rare.

Table 5.12

Auxiliary *be* errors

Family	child	wrong forms	correct forms	proportion of wrong forms
DZ1	A	1	240	0.00
	B	2	167	0.01
DZ2	A	1	202	0.00
	B	1	129	0.01
DZ3	A	1	28	0.03
	B	1	24	0.04
DZ4	A	0	171	0.00
	B	2	135	0.01
MZ1	A	0	169	0.00
	B	1	178	0.01
MZ2	A	7	75	0.09
	B	0	69	0.00
MZ3	A	1	132	0.01
	B	0	87	0.00
MZ4	A	0	91	0.00
	B	1	93	0.01

Thus a crucial prediction of the OI model, that children do not misapply *Infl* markings but simply omit them sometimes, is supported by the twins' data. This finding provides further support that the OI model is on the right track in characterizing children's errors as the optional application of tense and/or agreement rather than as mistakes due to incomplete learning.

5.4 Conclusion

Using the twin method, this study has demonstrated for the first time that there is a genetic component to individual differences in the use of tense and agreement by showing that MZ twins had a more similar course of development of these markers than DZ twins. MZ twins had higher correlations in the percent of obligatory tensing over time. MZ twins also had higher partial correlations with age partialed out of tensing over time, showing that their similarity comes not just from the fact that they are both getting older, but because their *Infl* systems are developing in synchrony. Furthermore, even with MLU partialed out of the development of tense/agreement markers, MZ twins were more similar than DZ twins. Regression coefficients between MZ twins' tensing over time with one twin's MLU partialed out were considerably higher than the same coefficients for DZs.

These findings support Wexler's (1996) conjecture that the development of *Infl* has a genetic basis. Wexler made this prediction for two reasons. First, the presumed genetic basis of the (Optional Infinitive related) impairment in SLI children. Second, the inexplicably long period of time normal children apparently take to master the use of tense/agreement markers. This second fact is an example of the "triggering problem" of Borer & Wexler (1987). That is, the data that children need to solve the problem is amply provided, and yet they fail to solve it for a protracted period. Wexler has more than once used the triggering problem to argue for the genetic basis of a linguistic ability (Borer & Wexler 1987, 1992; Wexler 1996b). However, this research represents the first time an argument about the genetic basis of syntax has been vindicated by behavior genetic research.

Chapter 6

Conclusion and Directions for Future Research

6.1 Summary of Findings

The purpose of this research was to find out whether language acquisition is heritable. It didn't have to be. Although it is widely agreed by now that adult language-related skills are heritable, nothing rests on the process of language acquisition itself being heritable. For one, there could be so little variation in the process that even if it were genetically specified, it would not be heritable. Of course, we know this is not the case; there is a lot of quantitative variation in language development, as has been pointed out in each chapter of this thesis and as anyone who has worked with children knows. Another possible state of affairs that would result in language development having no heritability is if even though the final state, adult language, is heritable, the process of getting there is driven by environmental differences. Some kids might learn faster than others because they have a better environment in some relevant respect. The results in this thesis argue against that possibility.

In Chapter 2, a study on vocabulary development was presented in which we saw that MZ twins were more similar than DZ twins in the age of reaching every milestone of productive vocabulary growth (the 10th word, the 25th word, the 50th word, the 75th word, the 100th word, and the 125th word) and more similar in their rate of word learning. They were also more similar in the strength of their vocabulary spurt, in the actual words they learned, and in the categories of words they learned. All of these effects were small but in several cases statistically significant by t-test, supporting the existence of genetic effects in vocabulary growth However, taking the results at face value for the moment, (i.e., assuming there are no problems with the methodology that was used to collect them), it appears that the rate and content of early vocabulary development has very little or no heritability. Using the DeFries-Fulker multiple regression method, h^2 was estimated at 10% or less for every measure in the vocabulary study and had high enough error that it was not reliably different from zero. Instead, vocabulary development had high and statistically significant shared environment.

In Chapter 3, results were presented from a study on first word combinations. This earliest of syntactic milestones was shown to be heritable. MZ twins were more similar than DZ twins both in the age of using their first productive multi-word phrases and in the rate of using them. This effect was significant by *t*-test and both measures also had high heritability, reliably above zero. In fact, h^2 was so high (greater than 1) in these analyses that c^2 had to be constrained to zero to estimate it. Taken at face value, these results indicate that the first glimmerings of syntax, combing words into rudimentary phrases, is heritable and that the effect of shared environment is negligible.

The focus of Chapter 4 was a study of the English past tense formation rule "add *-ed*," with the goal of using this rule as a case study of a grammatical rule. Studying overapplications of it allowed us to be sure that the rule is being applied productively, since the overregularized forms cannot have been learned from the child's input. The age at the use of the first overregularization did not show any heritability, and in fact did not provide very well-behaved data, with large variances for both MZ and DZ twins. A somewhat more stable measure, the rate of overregularization over a period of about 1 year, did show some signs of heritability. MZ twins had more similar rates than DZ twins, an effect which was significant by t-test and which produced a moderate h^2 of .44 (though not reliability higher than zero). The meaning of this result is ambiguous: it could mean learning a grammatical rule is heritable, or it could reflect the heritability of memory, since overregularizations occur when memory for a correct irregular past tense fails

Finally in Chapter 5 we presented a study with a very different methodology and a much smaller number of twins in order to be able to follow a developmental trend in excruciating detail. The goal was to find out whether the course of the Optional Infinitives stage, in which tense and/or agreement appear to be used only optionally by the child, is heritable. Despite the small number of twins and the occasionally small sample size from some transcripts, a definite trend emerged. The four pairs of MZ twins had more similar trajectories in achieving the consistent use of *Infl* than the four pairs of DZ twins. Even with MLU partialed out of the rate of using Infl over time, MZs were clearly more highly correlated than DZs, indicating for the first time that normal variation in the rate of using Infl is has a genetic basis (as opposed to the disordered state, which also appears to be genetically based).

6.2 Implications

Taking these findings without reservation for a moment, some clear trends have emerged. Early productive vocabulary appears to be driven largely by shared environmental rather than heritable sources, whereas the syntactic aspects of language we studied—word combinations, and the development of *Infl*—appear to have large heritable contributions. (The findings from the overregularization study are ambiguous at this point so I will leave them out of the discussion for the moment.) Perhaps a finding of low heritability in vocabulary development should not be so surprising after the publication of Reznick, Corley & Robinson 1997. Although they used a considerably different measure of early productive vocabulary, they found no heritability at 14 months, very little at 20 months, and a modest amount at 24 months. Since the twins in the vocabulary study reported here were typically between 14 and 20 months, finding about the same amount of heritability might be expected.

One possible explanation for the difference in heritability between vocabulary and word combinations is just age: heritability typically increases over time in any domain. Since children are older when they start putting words together than when they use them in isolation, the two possibilities are confounded. However, when behavior geneticists talk about heritability increasing over time, they typically mean over an entire life span, with peaks in old age. Such dramatic changes in heritability as those found between Chapter 1 and Chapter 2 in such a short period of time are unprecedented.

A more principled explanation of the contrast in heritability between syntactic development and vocabulary growth is that the difference reflects the underlying differences between vocabulary and syntax as processes. Vocabulary relies on memory for particular, explicit forms, while syntax relies on the implicit learning of rules. Vocabulary must be learned from the environment, and even the rate at which children learn words must ultimately be limited to the words they are exposed to. Syntax, on the other hand, cannot be learned from the environment; parents' efforts to teach syntax meet with resounding failure (Brown & Hanlon 1970; Marcus 1993) and parental speech has little effect on the character of child speech (Newport, Gleitman & Gleitman 1977). Thus the finding of high shared environment/low heritability for vocabulary and high heritability/low shared environment for word combinations is intuitively appealing. The difference further supports the notion that word learning and the combinatorial nature of syntax are different modules of language that are learned in different ways. However, more research will be needed to show that the two processes stem from different sets of genes. The same genes could be at work in both and simply interact with the environment in different ways.

6.3 Caveats

Each of these findings is novel and exciting, being the first attempt to establish heritability in language development and finding some strong effects and interesting trends. However, it may be too soon to take some of these results at face value since the experiments have some methodological flaws.

In vocabulary development, a finding of low heritability might be reasonable, but correlations of .97 for MZ twins and .92 for DZ twins are not. Even in Reznick, Corley & Robinson's (1997) study, where there was little heritability for productive vocabulary, correlations were much lower, in the range of .30 - .40 at 14 months and .60 - .70 at 20 and 24 months. Clearly the DZ and possibly the MZ correlations were inflated in the present study. One possibility is that there is a great deal of rater bias introduced by having the same rater for both twins in each pair. Although our efforts to quantify this bias in Chapter 2 did not reveal a lot of it, that analysis was based on only 10 pairs of twins who happened to have two raters. It could be that the very fact that there were two raters made them more accurate—either because they compared notes or because the parents are the kind of people who take their data collection responsibility seriously enough to make provisions to have records kept when they're not around.

Another possibility is that the first 100 or so words a child uses are often hard to understand, even for parents. Parents may unconsciously expect that once they hear one of their twins use a word, the other one will begin using it, too. They may take the slightest hint of the use of that word as legitimate in the co-twin. Basically, there is a lot of subjectivity in deciding what a 1-year-old says. This leaves a lot of room for error, and parents may be inclined to err on the side of increasing twin similarity for both MZ and DZ twins.

In the word combinations study, the opposite effect obtained: MZ correlations were more than twice as large as DZ correlations, yielding extremely high heritability. Although the explanation for that finding may lie in non-additive genetic effects, methodological flaws may again be to blame. Many of the same problems described in the vocabulary study apply here since he methodology was the same, although parental recording of word combinations is arguably more reliable since a word combination is easier to identify as such than a single word (for example, the intonational pattern of a multiple-word phrase in unmistakable, while the content of a single word is easy to mistake). Perhaps another effect is at work in this study: a contrast effect. Parents may be unconsciously trying to make their DZ twins less similar and their MZ twins more similar. However, failure to find any sort of contrast effect in the vocabulary study, plus the more objective nature of the task in the word combination study, make this possibility unlikely. For now the cause of the wildly large heritability and negative shared environment remains a problem to be dealt with in future research.

In the overregularization study, moderate heritability was found for rate. Since the rates of overregularization that parents provided us with are not wildly different from those in Marcus *et al* 1992, this may be a task relatively insulated from parent bias. Both overregularizations and irregular past tense forms are fairly salient and by the age they are producing them, children's pronunciation is typically much improved over the one-word stage. However, two additional problems plagued this study.

The first problem is in the overall tensing (not overregularization) result, where no heritability was found. It seems likely that parents are not reliable recorders of stem forms, which go into these counts, and with good reason. It is hard to explain the concept and importance of stem forms to an untrained parent, and it is often difficult to judge whether the context warranted a past tense, especially when records are kept in real time and there is no recording to check. However, the failure of parents to detect all the stem forms affects only the tensing rate, not the overregularization rate.

The second problem bears directly on the overregularization rate, and that is the "twin effect," or the effect of one twin using verbs that are overregularizations (to an adult) because he has heard his co-twin use them. The twin effect, if real, would raise apparent overregularization rates in both twins and would mean that some "overregularizations" from our point of view are correct forms from the child's point of view rather than true overapplications of the past tense rule. The biggest danger of the twin effect is that, since MZs were more likely to use the same words than DZs in the single word stage (a result reported in Chapter 2), MZs may have more of a twin effect in overregularization than DZs as well. Therefore
the twin effect warrants further investigation before the heritability results can be taken seriously.

In Chapter 5, we are finally free from the problems of parent report. However we were limited to only 8 pairs of twins. Although the effects we found appeared to be quite robust and in some cases even significant by *t*-test, it would be nice be able to test heritability models directly.

6.4 Goals of follow-up research

As these studies continue, it is clear from this discussion that two major issues need to be dealt with. The first is methodological. More follow-up work must be done to determine the validity of the parent-report method we've used and/or some new and more reliable method must be used instead. The second is conceptual. Although some attempt was made (in Chapters 2 and 3) to partial out motor development from linguistic development, clearly more careful data should have been gathered along with the language data on motor and non-linguistic cognitive development. Plans are underway to add these measures to the study as it continues, including a parent-administered non-linguistic test called the Parent Report of Children's Abilities (PARCA), developed and validated as part of the Twins' Early Development Study (TEDS) in the U.K. by Saudino *et al.* 1997.

This thesis is the only first step towards demonstrating and specifying the heritability of very early language development using spontaneous language production. As the first step it is exciting, but as the first step it represents only the beginning of the journey towards verifying the existence and extent of genetic variance in language acquisition.

References

Akerman, B. A., & Thomassen, P. A. (1991). Four-year follow-up of locomotor and language development in 34 twin pairs. *Acta Geneticae Medicae et Gemellologiae, 40,* 21-27.

Baker, M. (1988). *Incorporation.* Chicago: University of Chicago Press.

Bates, E. Bretherton, I. & Snyder L. (1988). *From first words to grammar.* Cambridge: Cambridge University Press.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In Fletcher, P. & MacWhinney, B. (Eds), *The handbook of child language.*

Bayley, N. (1969). *Manual for the Bayley Scales of Infant Development.* New York: Psychological Corporation.

Berko, J. (1958). The children's learning of English morphology. *Word, 14,* 150-177.

Bloom, L. (1973). *One word at a time: The use of single-word utterances before syntax.* The Hague: Morton.

Bloom, L., Lightbown, P, & Hood, L. (1976). Structure and variation in child language. *Monographs of the Society for Research in Child Development, 40,* serial no. 160.

Bobaljik, J. (1996). *Morphosyntax: The syntax of verbal inflection.* Ph.D. Thesis, MIT. Distributed by MITWPL.

Borer, H., and Wexler, K. (1987). The Maturation of Syntax. In Thomas Roeper and Edwin Williams (eds.), *Parameter Setting,* pp. 123-172. D. Reidel Publishing Company.

Borer, H., and Wexler, K. (1992). Bi-unique Relations and the Maturation of Grammatical Principles. *Natural Language and Linguistic Theory, 10,* 147-189. The Netherlands: Kluwer Academic Publishers.

Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science,* 250, pp. 223-228.

Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science,* 212, 1055-1059.

Brain, M.D.S. (1963). The ontogeny of English phrase structure: The first phrase. *Language, 39,* 1-14.

Braine, M.D.S. (1976). Children's First Word Combinations. *Monographs of the Society for Research in Child Development, 41,* serial no. 164.

Bromberg, H.S., & Wexler, K. Null-subjects in wh-questions. In C. T. Schutze, J.B. Ganger, & K. Broihier (Eds.), 1995 *Papers on language processing and acquisition: MIT Working Papers in Linguistics, 26,* 221-247.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Brown, R. & Fraser, C. (1963). The acquisition of syntax. In C. N. Cofer and B. Musgrave (Eds), *Verbal Behavior and Learning: Problems and Processes.* New York: McGraw-Hill, pp. 158-201.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the Development of Language,* pp. 11-53. New York: Academic Press.

Cardon, L. R., Smith, S. D., Fulker, D. W., Kimberling, W. J., Pennington, B. F., DeFries, J. C. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science,* 266, 276.

Cazden, C. B. (1968). The acquisition of noun and verb inflection. *Child Development,* 39, 433-448.

Chomsky, N. (1957). *Syntactic structures.* The Hague.

Chomsky, N. (1991). Some notes on economy of derivation and representation. Reprinted Chomsky (1995), pp. 129-166.

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

Chubrich, R. (1971). *Selected phonological, morphological, and syntactic measures for identical and non-identical same-sex twins.* Ph.D. Dissertation, SUNY Buffalo.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd ed.* Lawrence Erlbaum: Hillsadale, NJ.

Conway, D., Lytton, H., & Pysh, F. (1980). Twin-singleton language differences. *Canad. J. Behav. Sci,* 12, 264-271.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences,* 14

Davis, E. A. (1937). The development of linguistic skill in twins, singletons, and sibs, and only children from 5-10. *University of Minnesota Institute of Child Welfare, Monograph 14.*

Day, E. J. (1932) The development of language in twins. 1. A comparison of twins and single children. 2. The development of twins: their resemblances and differences. *Child Development,* 3, 179-199, 298-316.

DeFries, J. C. & Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavior Genetics,* 15, 467-473.

DeFries, J. C. & Fulker, D. W. (1988). Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellologiae, 37,* 205-216.

Dodd, B. & McEvoy, S. (1994). Twin language or phonological disorder? *Journal of Child Language,* 21, 273-289.

Dromi, E. (1987). *Early lexical development.* Cambridge: Cambridge University Press.

Dunn, L. M. (1965). *Expanded Manual for the Peabody Picture Vocabulary Test.* Circle Pines, MN: American Guidance Service.

Eaves, L. J. (1982). The utility of twins. In V. Anderson *et al.* (Eds.), *Genetic basis of the epilepses.* New York: Raven Press.

Emonds, J. E. (1976). *A transformational approach to English syntax.* Academic Press.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., and Thal, D. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development, 59,* serial no. 242.

Fischer, K. M. (1973). *A comparison of the similarity in language skills of identical and fraternal twin pairs.* Ph.D. dissertation, University of Pennsylvania.

Foch, T. T., & Plomin, R. (1980). Specific cognitive abilities in 5- to 12-year old twins. *Behavior Genetics, 10,* 507-520.

Foster, R., Giddon, J., & Stark, J. (1972). *Manual for the Assessment of Children's Language Comprehension.* Palo Alto, CA: Consulting Psychologists Press.

Gallaway, C. & Richards, J.B. (Eds.) (1994). *Input and interaction in language acquisition.* Cambridge: Cambridge University Press.

Galton, F. (1876). The history of twins as a criterion of the relative powers of nature and nurture. *Royal Anthropological Institute of Great Britain and Ireland Journal,* 6, 353-357.

Gathercole, V. (1979). *Birdies like birdseed the bester than buns: A study of relational comparatives and their acquisition.* Unpublished Ph.D. dissertation, University of Kansas.

Geschwind, N. (1983) Genetics: Fate, chance, and environmental control. In C. Ludlow and J. Cooper (Eds), *Genetic aspects of speech and language disorders,* pp. 21-33.New York: Academic Press.

Gleason, J.B. (1980). The acquisition of social speech and politeness formulae. In H. A. R. Giles (Ed), *Language: Social psychological perspectives.* Oxford: Pergamon.

Goldin-Meadow, S., & Mylander, C. (1984). Gestural communication in deaf children: The non-effects of parental input on early language development. *Monographs of Society for Research in Child Development* 49, serial no. 207.

Goldfield, B. & Reznick, J. S. (1990). Early lexical acquisition: Rate content, and the vocabulary spurt. *Journal of Child Language 17,* 171-183.

Goldsmith, H. H. (1991). A zygosity questionnaire for young twins: A research note. *Behavior Genetics, 21,* 257-269.

Griffiths, A.T.F., Miller, J.H., Suzuki, D. T., Lewontin, R.C., & Gelbart, W.M. (1993). *An introduction to genetic analysis (5th edition).* New York: W. H. Freeman & Co.

Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J . Keyser (Eds), *The View from Building 20.* Cambridge, Mass: MIT Press.
Harris, T. & Wexler, K. (1996). The optional-infinitive stage in child English: Evidence from negation. In H. Clahson (Ed.), *Generative perspectives on Language Acquisition.* Amsterdam: John Benjamins.
Hay, D. A., Prior, M., Collett, S., Williams, M. (1987). Speech and language development in preschool twins. *Acta Geneticae Medicae et Gemellologiae, 26*, 213-223.
Heath, A. C. & Martin, N. G.. (1986). Detecting the effects of genotype x environment interaction on personality and symptoms of anxiety and depression. *Behavior Genetics, 16,* 622.
Hedrick, D.L., Prather, E.M., & Tobin, A. R. (1975). *Sequenced inventory of communication development.* Seattle: University of Washington Press.
Jonas, D. (1995a). On the acquisition of verb syntax in child Faroese. *MIT Working Papers in Linguistics*, 26, 265-280.
Jonas, D. (1995b). *Clause structure and verb syntax in Scandinavian and English.* Unpublished doctoral dissertation, Harvard University.
Koch, H.L. (1966). *Twins and twin relations.* Chicago University Press: Chicago.
Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16: 589-600.
Langacker, R.W. (1987). *Foundations of cognitive grammar, Volume I: Theoretical prerequisites.* Stanford, CA: Stanford University Press.
Lee, L. (1971). *Northwestern Syntactic Screening Test.* Evanston, IL: Northwestern University Press.
Lenneberg, E. H. (1967). *Biological Foundations of Language.* New York: John Wiley & Sons.
Lieven, E.V.M., Pine, J.M., & Dresner-Barnes, H. (1992). Individual differences in early vocabulary development: redefining the referential-expressive distinction. *Journal of Child Language*, 19, 287-310.
Loehlin, J.C., & Nichols, R.C. (1976). *Heredity, environment, and personality.* Austin: University of Texas Press.
Luria, A. R, & Yudovich, F. (1959). *Speech and the development of mental processes in the child.* Staples, London.
Lykken, D. T., McGue, M., Tellegen, A., & Bouchard Jr., T. J. (1992). Emergenesis. *American Psychologist*, 47, 1565-1577.
Lytton, H. (1980). Parent-child interaction. The socialization process observed in twin and singleton families. Plenum, NY.
MacWhinney, B. & Snow, C. (1990). The Child Language Data Exchange System. *Journal of Child Language*, 17, 457-472.
Malmstrom, P. M., & Silva, M. N. (1986). Twin talk: manifestations of twins status in the speech of toddlers. Jo*urnal of Child Language*, 13, 293-304.
Marcus, G. (1993). Negative evidence in language acquisition. *Cognition*, 46: 53-85.
Marcus, G.F., Pinker, S., Ullman, M., Hollander, M., Rosen, T.J., Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57,* serial no. 228.
Martin, N.G., Eaves, L.J., Kearsey, M.J., & Davies, P. (1978). The power of the classical twin study. *Heredity*, 40, 97-116.
Mather, P. L. & Black, K. N. (1984). Hereditary and environmental influences on preschool twins' language skills. *Developmental Psychology, 20*, 303-308.
Mather, K. & Jinks, J.L. (1982). Biometrical *genetics: The study of continuous variation* (3rd ed.). London: Chapman and Hall.
McCarthy, J.J. (1954). Language development in children. In Leonard Carmichael (Ed), *Manual of Child Psychology,* 2nd ed. New York: Wiley, pp 492-630.
McCarthy, J.J & Kirk, S. A. (1961). *The Illinois Test of Psycholinguistic Abilities* (Experimental Edition). Urbana: IL: Institute for Research in Exceptional Children.

McEvoy, S.& Dodd, B. (1992). The communication abilities of 2- to 4-year old twins. *European Journal of Disorders of Communication*, 27, 73-87.
Mehrabian, A. (1970). Measures of vocabulary and grammatical skills for children up to age six. *Developmental Psychology, 2,* 439-446.
Miller, J.F. & Chapman, R. S. (1991). *Systematic analysis of language transcripts.* Madison: University of Wisconsin Language Analysis Lab.
Mittler, P. (1969). Genetic aspects of psycholinguistic abilities. *Journal of Child Psychology and Psychiatry, 10,* 165-176.
Mittler, P. (1970) Biological and social aspects of language in twins. *Developmental Medicine and Child Neurology,* 12, 741-757.
Morley, M. (1965). *The Development and Disorders of Speech in Children.* Edinburgh: E&S Livingstone.
Munsinger & Douglass (1976). The syntactic abilities of identical twins, fraternal, twins, and their siblings. *Child Development, 47,* 40-50.
National Society for the Study of Education. (1929). Preschool and parental education. *Yearbook of the National Society of Studies in Education* 28, 495-568. Cited in McCarthy 1954.
Neale, M. C. & Maes, H. H. M. (in prep). *Methodology for genetic studies of twins and families.* Dordrecht: Kluwer Academic Publishers.
Neale, M. C. & Stevenson, J. (1989). Rater bias in the EASI temperament scales: A twin study. *Journal of Personality and Social Psychology*, pp. 446-455.
Nelson, K. (1973). Structure and Strategy in Learning to Talk. *Monographs of the Society for Research in Child Development, 38,* Serial No. 149.
Nelson, K. (1981). Individual differences in language development: Implications for development and language. *Developmental Psychology*, 17, 170-187.
Neils, J. & Aram, D. M. (1986). Family history of children with developmental language disorders. *Perceptual and Motor Skills Part I*, 63(2), 655-658.
Newport, Elissa L., H. Gleitman and L. R. Gleitman. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. Snow and C. Ferguson (eds.) *Talking to children.* Cambridge: Cambridge University Press.
Nichols, , R. C. (1978). Twin studies of ability, personality, and interests. *Homo*, 29, 158-173.
Noble, E. (1980). *Having twins: a parent's guide to pregnancy, birth, and early childhood.* Boston: Houghton Mifflin.
Pennington, B. (1990). The genetics of dyslexia. *Journal of Child Psychology and Psychiatry*, 31, 193-201.
Phelps, J. A., Davis, J. O., Schartz, K. M. (1997). Nature, nurture, and twin research strategies. *Current Directions in Psychological Science*, 6, 117-121.
Pierce, A. (1992). *Language acquisition and syntactic theory.* Dordrecht, The Netherlands: Kluwer.
Pine, J.M., Lieven, E.V.M., Rowland, C. (1996). Observational and checklist measures of vocabulary composition: what do they mean? *Journal of Child Language, 23,* 573-589.
Pinker, S. (1984). *Language Learnability and Language Development.* Cambridge, MA: Harvard University Press.
Plomin, R., DeFries, J.C., & McClearn, G. E. (1990). *Behavior Genetics: A primer* (2nd ed). New York: W. H. Freeman.
Plomin, R., Emde, R. N., Braungart, J. M., Campos, J., Corley, R., Fulker, D. W., Kagan, J., Reznick, J. S., Robinson, J., Zahn-Waxler, C., & DeFries, J. C. (1993). Genetic change and continuity from fourteen to twenty months: the MacArthur Longitudinal Twin Study. *Child Development, 64,* 1354-1376.
Poeppel, D. & Wexler, K. (1993). The full competence hypothesis of clausal structure in early German. *Language, 69,* 1-33.
Pollock, J.-Y. (1989). Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry*, 20, 365-424.
Prescott, G. A., Balow, I. H., Hogan, T. P., & Farr, R. C. (1986). *Metropolitan Achievement Tests.* New York: The Psychological Corporation.

Record, R. G., McKeown, T., & Edwards, J. H. (1970). An investigation of the difference in measured intelligence between twins and single births. *Annals of Human Genetics*, 34, 11-20.
Reznick, J. S. (1990). Visual preference as a test of infant word comprehension. *Applied Psycholinguistics*, 11:145-165.
Reznick, J.S. (1996). Intelligence, nature, and nurture in young twins. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Intelligence, Heredity, and Environment*. Cambridge University Press.
Reznick, J.S., & Goldfield, B. A. (1994). Diary vs. representative checklist assessment of productive vocabulary. *Journal of Child Language*, 21, 465-472.
Reznick, J. S., Corley, R., & Robinson, J. (1997). A longitudinal twin study of intelligence in the second year. *Monographs of the Society for Research in Child Development, 62,* serial no. 249
Rice, M. L., Rice, K. J., & Wexler, K. (In press) "Family histories of language impaired children in the Extended Optional Infinitive stage of development", *Journal of Speech and Hearing Research.*
Rice, M. & Wexler, K. (1996) Toward tense as a clinical marker of Specific Language Impairment in English-speaking children. *Journal of Speech and Hearing Research*, 39, 1239-1257.
Rice, M., Wexler, K., & Cleave, P. L. (1995). Specific Language Impairment as a period of extended Optional Infinitive. *Journal of Speech and Hearing Research, 38,* 850-863.
Rice, M., Wexler, K., & Hershberger, S. (In press). Tense over time: The longitudinal course of tense acquisition in children with Specific Language Impairment. *Journal of Speech and Hearing Research.*
Saudino, K., Dale, P. S., Oliver, B., Petrill, S. A., Richardson, V., Rutter, M., Simonoff, E., Stevenson, J., & Plomin, R. (1997). *British Journal of Developmental Psychology.*
Savic, S. (1980). *How twins learn to talk.* New York: Academic Press.
Scarr, S. (1968). Environmental bias in twin studies. *Eugenics Quarterly*, 15, 34-40.
Scarr, S. (1969). Effects of birth weight on later intelligence. *Social Biology*, 16, 249-256.
Scarr, S. & Carter-Saltzmann, L. (1979). Twin method: defense of a critical assumption. *Behavior Genetics*, 9, 527-542.
Schutze, C. T., & Wexler, K. (1996). Subject case licensing and English root infinitives. In A. Stringfellow, D. Cahana-Amitay, E. Hughes, & A. Zukowski (Eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development,* 670-681.
Segal, N. (1985). Monozygotic and dizygotic twins: a comparative analysis of mental ability profiles. *Child Development, 56,* 1051-1058.
Siegel, S. & Castellan Jr., N. J. (1988). *Nonparametric statistics for the behavioral sciences.* Boston, MA: McGraw-Hill.
Snedecor, G.W. & Cochran, W.G. (1989). *Statistical Methods*, 8th ed. Ames: Iowa State University Press.
Snow, C. E. & Ferguson (Eds.). (1977). *Talking to children: Language input and acquisition.* Cambridge: Cambridge University Press.
Snyder, W. & Bar-Shalom, E. (1997). Word order, finiteness, and negation in early child Russian. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development*, pp. 717-725.
Snyder, W., & Stromswold, K. (1997). The structure and acquisition of English dative constructions. *Linguistic Inquiry, 28,* 281-317.
Spitz, E., Moutier, R., Reed, T., Busnel, M. C., Marchaland, C., Roubertoux, P. L., & Carlier, M. (1996). Comparative diagnoses of twin zygosity by SSLP variant analysis, questionnaire, and dermatoglyphic analysis. *Behavior Genetics, 26,* 55-63.
Stevenson, J., Graham, P., Fredman, G., McLoughlin, V. (1987). A twin study of genetic influences on reading and spelling ability and disability. *Journal of Child Psychology and Psychiatry*, 28:229-247.

Stevenson, J., Pennington, B. F., Gilger, J. W., DeFries, J. C., Gillis, J. J. (1993). Hyperactivity and spelling disability: testing for a shared genetic aetiology. *Journal of Child Psychology and Psychiatry*, 34: 1137-1152.
Stromswold, K. (1990). *Learnability and acquisition of auxiliaries*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
Stromswold, K. (under review). The heritability of language: A review of twin and adoption studies. Ms., Rutgers University.
Tallal, P., Ross, R., & Curtiss, S. (1989). Familial aggregation in specific language impairment. *Journal of Speech and Hearing Research*, 54, 167-173.
Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale: Manual for the Third Revision, Form L-M*. Boston: Houghton-Mifflin.
Tomasello, M., Mannle, S. & Kruger, A. (1986). Linguistic environment of 1- to 2-year-old twins. *Developmental Psychology*, 22, 169-176.
Thompson, L. A., Detterman, D. K., & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: genetic overlap but environmental differences. *Psychological Science, 2*,158-165.
Tomblin, J. B. (1989). Familial concentration of developmental language impairment. *Journal of Speech and Hearing Disorders*, 54, 287-295.
Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A Neural Dissociation within Language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9, 266-276.
Wahlsten, D. (1994). The intelligence of heritability. *Canadian Psychology*, 35:3, 244-260.
Warren-Leubecker, A. (1982). *Sex differences in speech to children*. Unpublished master's thesis, Georgia Institute of Technology.
Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corporation.
Wexler, K. (1990). Recent studies in the development of inflection. Paper presented at the symposium on current reserch in language acquisition at the annual meeting of the Society for Cognitive Science, MIT.
Wexler, K. (1994) "Optional infinitives, head movement and the economy of derivations", in D. Lightfoot & N. Hornstein (Eds.), *Verb movement*. Cambridge, England, Cambridge University Press, pp. 305-350.
Wexler, K. (1996a). The development of inflection in a biologically based theory of language acquisition. In M. Rice (ed), *Towards a genetics of language*, pp. 113-144.
Wexler, K. (1996b). Maturation and growth of grammar. In W. C. Ritchie & T. K. Bhatia, (Eds.), *Handbook of Language Acquisition*.
Wexler, K. (In press). Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* 106.
Wilson, R. (1975). Twins: Patterns of cognitive development as measured on the Wechsler Preschool and Primary Scale of Intelligence. *Developmental Psychology*, 11, 126-134.
Wilson, R. (1985). Birthweight and IQ. *Developmental Psychology, 21*, 795-805.
Zazzo, R. (1960). *Les jumeaux: Le couple et la personne*. Paris: P.U.F.

<u>Appendix A</u>

Phone protocol used to recruit subjects

MIT Twins Study

Answer with "hello"

Thanks for calling the MIT Twins Study.

I am ____________

and your name is?

And you have twins? Are they both the same sex? How old are they?

(If not same sex, go to option 1)

Option 1 We aren't using <population> twins right now, but may we take your name and address for future studies? (Take name, address, and phone number, as well as names of twins and date of birth.) Thank your for your interest.

(If younger than 8 months, go to option 2)

Option 2: We aren't starting the study until the children are ready to say their first words, which is around 10 to 14 months, but we would like to call you back at that time. May we take your name and address? (Also get phone number, as well as names of twins and date of birth.)

Where are you calling from?

(If long distance, take phone number and offer to return call.)

Are your twins identical or fraternal?

How do you know?

Are your children only exposed to English?

(If not, go to Option 1)

Have your twins started to talk yet?

(Get as much info as possible about language production in order to figure out where to start with checklists. If they have more than 5 words each, only tell them about Checklist #2 when you get to that part of the call.)

I need to ask you some more personal questions now, and you don't have to answer if you don't feel comfortable giving out the information. Is that okay?

Were your twins born full-term?

(If not, how long was gestation?) (37 weeks is average . . . if they were less than 34 weeks start to worry)

What were their birth weights?

(If they were under 3 pounds -OR- if they are identical and their weights have a discrepancy of more than 1 pound, maybe, go to option 1 to reject them.)

What is their exact date of birth?

How much time did they have to spend in the hospital after they were born?

(If it was more than 1 week, ask why.)

IF FRATERNAL: Were the twins spontaneous? (Were any fertility treatments used?)

Have they been diagnosed with any speech, language, or hearing problems?

Do they have a history of ear infections? (If they are chronic, exlude)

What about physical or neurological problems?

(If major impairment, go to option 1)

Has anyone is your (or the father's) immediate family been diagnosed with speech or language problems?

Okay, we're done with those questions.

For our records, where did you hear about our study?

Checklist Description:

What we're doing is trying to find out what aspects of language have a genetic basis and what aspects have to be learned from parents. So, we're doing a long-term study of language development, starting from your twins' first words and continuing until they know about a hundred words. This could take just a few months or up to 6 or 7 months, depending on how quickly they develop.

The way we keep track of their development is by relying on you. We ask you to keep a daily journal of the different words your twins say every day. We give you organized sheets of paper to write the words down on, with space for each twin. You would start a new sheet each day and then send them back to us in batches of 7 each week in stamped envelopes which we provide.

At first, this will be pretty easy; the twins are probably only saying a few words each day. When they've gotten up to 30 or 40 words in their vocabulary, it will get harder, and you can cut down to just keeping this diary for an hour a day as many days per week as possible. The important thing is to be as consistent as possible and be sure you're writing down the right words for the right twin.

I should warn you that although a lot of parents find it fun to follow their twins' language development so closely, it is a bit of a commitment, since you'll be keeping this journal for us for several months.

We don't have too much to offer you in the way of compensation for participating, but we do offer either $50 or a zygosity test, costs $120, after we receive the first ten weeks. We also give you a copy of the book

The Language Instinct by Steven Pinker, who is in charge of this study. It's a very readable book which just explains generally why we think there is a genetic basis to language. Finally, at the end of the study, we'd send you a pair of appropriately sized MIT Twins Study t-shirts for your twins and you would also receive a profile of your children's language development and the results of the study when the study is over, which will be at least a year from now.

Does that sound like something you'd like to take part in?

If YES, get name, address, phone number, twins names and exact date of birth. Ask when is a good time to reach them on the phone. Tell them: We will send out a packet to you <on the next mailing> , and then call you to go over the instructions. The packet will include the Language Instinct, some permission forms and background questionnaires to send back in, and three weeks worth of diary pages to fill out. We'll call soon to make sure you understand the instructions, and we'll call every three weeks after that to see how things are going. We'll also send new checklists every three weeks. You should always feel free to call us with questions whenever you have them (make sure they have phone number).

If NO, thank for their interest. Get info for database if possible.

Appendix B

Sample of the form used by parents, instructions given to parents, and an example form given to parents in the vocabulary study

MIT Twins Study: *First Words* Checklist   Date____/____, 1998

Family Name ____________________

*Name*______________________

WORD or PHRASE	Comments
	Please answer the following:
	How is child feeling today?
	Who is keeping this list today?

*Name*______________________

WORD or PHRASE	
	How is child feeling today?
	Who is keeping this list today (if different from above)?

Instructions for First Words checklist

Start keeping these lists on the first day that either twin says his or her first word.

Write down each word your twins say each day in the space provided. Use a new sheet each day. If one child says the same word more than once per day, you only have to write it down the first time. You may not need to use all the spaces provided. If you run out of space, continue on another page.

If you cannot tell whether something is a word, make your best guess. Some questions to base this guess on are: Does the child consistently use this "word" to refer to a particular object and no other objects? Does the child say the word repeatedly and become frustrated if s/he is not understood? If the answers are yes, then the phrase is probably a word. If you are still unsure, make a note to that effect in the *comments* space.

It is very important to keep each child's word list separate. The page is divided so that one twin's word list can go on top, and the other's on bottom. We recommend consistently assigning one twin to the top list and the other to the bottom to avoid confusion from day to day.

Use the comment space to explain what words refer to if the meaning is not clear. Many children have idiosyncratic words that are not recognizable by anyone but their parents. For example, a child may consistently use "waa" to refer to water.

Also use the comment space to note if a word is a "repetition" of something a child just heard. If a child immediately repeats a word that another person says, including a sibling, write the word down but note that it is a repetition.

Some of your children's words may be two-word combinations, such as "no more", "all gone", "bye-bye", etc. As long as these words are used as a unit, they should count as one word. If you are in doubt over whether a phrase counts as one or two words, note that in the space provided beside the word.

Finally there is a space for **two questions you should answer every day**: "How is the child feeling?" and "Who is keeping this list today?"

See the attached example sheet for more details. If you have any questions or comments, feel free to call us at (617)253-8408 or email us at jganger@psyche.mit.edu.

CHECKLISTS SHOULD BE MAILED BACK TO US IN SETS OF SEVEN EACH WEEK IN THE ENVELOPES PROVIDED. WE WILL SEND OUT NEW PACKETS OF CHECKLISTS AS NEEDED.

MIT Twins Study: *First Words* Checklist   Date 2 / 2 , 1998

Family Name ____________________

Name Dan

WORD or PHRASE	Comments
daddy	repeat of Dave
no	
all gone	
juice	
night-night	
	Please answer the following:
	How is child feeling today? getting over a cold
	Who is keeping this list today? mom

Name Dave

WORD or PHRASE	
daddy	
peek-a-boo	
apple	
nana	banana
bath	
duck	
	How is child feeling today? fine
	Who is keeping this list today (if different from above)?

<u>Appendix C</u>

Sample of the instructions given to parents and an example form given to parents in the word combination study (the form used by parents in the same as in Appendix B)

Instructions forMultiple-Word Phrases checklist

Start keeping these lists on the first day that either twin says his or her first multpile-word phrase. We consider a multiple-word phrase to be any combination of two or more words used as a unit. Examples include "bad doggy", "more juice", "bye bye daddy", "up please", "mommy go bye bye".

Write down each multiple-word phrase your twins say each day in the space provided. If one child says the same phrase more than once per day, you only have to write it down the first time. You may not need to use all the spaces provided. If you run out of space, continue on another page.

If you cannot tell whether something is a multiple-word phrase, write it down anyway, but mention your concern in the *comments* space.

It is very important to keep each child's phrase list separate. The page is divided so that one twin's phrase list can go on top, and the other's on bottom. We recommend consistently assigning one twin to the top list and the other to the bottom to avoid confusion from day to day.

Use the comment space to explain what phrases refer to if the meaning is not clear. Many children have idiosyncratic phrases that are not recognizable by anyone but their parents. For example, a child may consistently use "mo waa" to refer to more water.

Also use the comment space to note if a phrase is a "repetition" of something a child just heard. If a child immediately repeats a phrase that another person says, including a sibling, write the phrase down but note that it is a repetition.

Some of your children's multiple-word phrases may be two-word combinations that you think of as one word, such as "no more", "all gone", "bye-bye", etc. Write those phrases down as well. If you are in doubt over whether a phrase counts as one word or a true combination, note that in the space provided beside the word.

Finally, there is a space for **two questions you should answer every day**: "How is the child feeling?" and "Who is keeping this list today?"

See the attached example sheet for more details. If you have any questions or comments, feel free to call us at (617)253-8408 or email us at jganger@psyche.mit.edu.

CHECKLISTS SHOULD BE MAILED BACK TO US IN SETS OF SEVEN EACH WEEK IN THE ENVELOPES PROVIDED. WE WILL SEND OUT NEW PACKETS OF CHECKLISTS AS NEEDED.

MIT Twins Study: *First Words* Checklist Date 10 / 28 , 1998

Family Name ____________________

Name Michael

WORD or PHRASE	Comments
I do it	
more juice	repeat of mom
da car	daddy's car
all gone	
	Please answer the following:
	How is child feeling today? getting over a cold
	Who is keeping this list today? mom

Name Matthew

WORD or PHRASE	
da car	daddy's car--repeat of Michael
no more	
bye bye mommy	
wan juice	I want juice
me do	I do it
	How is child feeling today? fine
	Who is keeping this list today (if different from above)?

Appendix D

Sample of the form given to parents, instructions given to parents and an example form given to parents in the overregularizaton study

MIT Twins Study: *Hour-a-Day* Checklist Date____/____, 1998

Family Name ____________________

Please write in words--see instructions for details

*Name*______________________

Past tense (e.g., walked, sang, goed, fall)	Comments	Daily comments:
		Who is keeping this list today?
		How is child feeling today?

Comments or interesting sentences "Mommy told me put it away"

*Name*______________________

Past tense (e.g., walked, sang, goed, fall)	Comments	Daily comments::
		Who is keeping this list today?
		How is child feeling today?

Comments or interesting sentences ______________________

Instructions for *Past Tense Verbs* checklist

General Instructions

Start a new checklist each day, noting the date at the top.

It is very important to keep each child's records separate. The page is divided so that one twin's word list can go on top, and the other's on bottom. We recommend consistently assigning one twin to the top list and the other to the bottom to avoid confusion from day to day.

Keep in mind that just because your children are twins doesn't mean they will develop at exactly the same rate, whether they are identical or fraternal.

You should spend about an hour each day keeping this sheet nearby and listening for the things you need to write down. Of course, we expect that you will have to miss some days or that you may not be able to spend an entire hour on some days. However, consistency is important for this study, so we ask you to complete the checklists on as many days as you can. Furthermore, if you wish, you may keep track all day long, or write down any appropriate forms you notice throughout the day in addition to your one hour of concerted attention.

Checklists should be mailed back to us in sets of seven each week in the envelopes provided. We will send out new packets of checklists as needed. You are welcome to make copies of these checklists for your own records before you send them back to our office. When the study is completed, we will send all participants a more condensed version of their twins' development rather than sending back each checklist.

Any time a child's utterance is an immediate repetition of something you or someone else has just said, note this fact by writing "R" beside the word or phrase. Also place an "R" beside a word or phrase if the child immediately repeats him or herself without interruption (see sample checklist for examples).

On the other hand, if a child says the same thing more than once during the session and this phrase is **separated by other conversation**, make a tally mark beside the word or phrase for each additional time that it is said to indicate that the phrase was not just repeated (see sample checklist for examples). If you are in doubt as to whether something is a repetition or a tally, use R.

Instructions for this checklist: Please read carefully

The goal of this checklist is to record all of your twins' **past tense verb** uses and attempts at using them. You should use the first column (marked "Past tense verbs") to do this. The **second column** (marked "Comments") can be used to note any questions or comments that come up about particular forms you write down. The **third column** contains two questions you should answer every day: how is the child feeling, and who is keeping the list.

The past tense is usually formed by adding -ed to a verb, as in

walk--walked.

But some verbs have irregular, or exceptional past tense forms, for example

sing--sang
go--went
is--was

You should write down whenever you hear a verb that *is* or *should be* a past tense form.
For example, a child might say

"Yesterday <u>we went</u> to the library."

or

"Yesterday <u>we go</u> to the library."

or

"Yesterday we goed to the library."

All of these cases count as past tense forms for the checklists. We are interested in all of the forms, correct and incorrect, that your children produce. Write down both the subject and verb of the sentence and any other words you think are necessary. See the attached Sample Checklist for more examples.

If the child does not use the correct form, it can be hard to tell he/she really meant to use the past tense, as in, for example:

"I tell Daddy."

This could mean that the child just told Daddy something, or that s/he wants to tell Daddy something, etc. In these cases you have to use your own judgement from the context of the conversation. If you are not sure, jot down a note to this effect in the Comment space provided.

There is also space on the bottom of each half of the sheet for "**comments or interesting sentences**". If you want to make any comments about the entire session please put them here. Or, if you hear a sentence or phrase you think is interesting for some reason, write it down in that space.

See the attached example sheet for more details. If you have any questions or comments, feel free to call us at (617)253-8408 or e-mail us at jganger@psyche.mit.edu.

MIT Twins Study: *Hour-a-Day* Checklist Date 2 / 2 , 1998

Family Name ____________________

Please write in words--see instructions for details

Name Heidi

Past tense (e.g., walked, sang, goed, fall)	Comments	Daily comments:
we climb		Who is keeping this list today?
she go home	not sure if in past	
ball dropped		mother
she sang		
I picked it up	I I I	How is child feeling today? fine
my bumped it		

Comments or interesting sentences "Mommy told me put it away"

Name Holly

Past tense (e.g., walked, sang, goed, fall)	Comments	Daily comments::
Kitty jumped		Who is keeping this list today?
I fall down	R	
I goed		mother
Heidi tooked it		
		How is child feeling today? a little sick

Comments or interesting sentences

APPENDIX E

Raw data for 8 the pairs of twins (4 MZ, 4 DZ) in Chapter 5: Percent correct tensing overall and for each tense marker

MZ1A Session	age in months	average % correct	# correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	28.83	0.91	10	11	0.00	0	1	1.00	3	3	1.00	3	3	1.00	3	3	n/a	0	0	n/a	0	0	1.00	1	1
2	29.30	0.58	7	12	n/a	0	0	0.00	0	1	0.67	6	9	1.00	1	1	n/a	0	0	n/a	0	0	0.00	0	1
3	29.77	0.58	7	12	0.33	1	3	0.00	0	1	0.71	5	7	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1
4	31.00	0.50	1	2	n/a	0	0	n/a	0	0	0.00	0	1	1.00	1	1	n/a	0	0	n/a	0	0	n/a	0	0
5	31.90	0.50	1	2	n/a	0	0	n/a	0	0	0.00	0	1	1.00	1	1	n/a	0	0	n/a	0	0	n/a	0	0
6	32.37	0.48	34	71	0.25	2	8	0.17	1	6	0.49	18	37	1.00	1	1	0.17	1	6	n/a	0	0	0.85	11	13
7	32.83	0.50	21	42	n/a	0	0	0.00	0	2	0.47	17	36	1.00	1	1	n/a	0	0	1.00	1	1	1.00	2	2
8	33.30	0.39	9	23	0.00	0	2	0.50	3	6	0.33	4	12	n/a	0	0	0.00	0	1	n/a	0	0	1.00	2	2
9	33.87	0.64	39	61	0.60	3	5	0.50	7	14	0.71	24	34	n/a	0	0	0.00	0	1	n/a	0	0	0.71	5	7
10	34.57	0.62	40	64	0.36	4	11	1.00	1	1	0.66	21	32	1.00	3	3	0.00	0	3	1.00	3	3	0.73	8	11
11	35.03	0.53	34	64	0.60	3	5	0.20	3	15	0.67	16	24	1.00	3	3	0.67	2	3	0.71	5	7	0.29	2	7
12	37.80	0.92	33	36	1.00	6	6	1.00	9	9	0.88	14	16	1.00	1	1	1.00	2	2	1.00	1	1	0.00	0	1
13	38.27	0.74	31	42	1.00	7	7	0.60	6	10	0.71	10	14	1.00	1	1	0.25	1	4	1.00	2	2	1.00	4	4
14	38.73	0.83	20	24	1.00	4	4	1.00	2	2	0.92	12	13	n/a	0	0	1.00	1	1	n/a	0	0	0.25	1	4
15	39.17	0.79	27	34	0.86	6	7	0.40	2	5	1.00	15	15	1.00	2	2	n/a	0	0	1.00	1	1	0.25	1	4
16	40.10	0.94	30	32	1.00	5	5	1.00	4	4	0.90	19	21	n/a	0	0	n/a	0	0	1.00	1	1	1.00	1	1
17	40.60	0.85	28	33	1.00	4	4	0.50	1	2	1.00	17	17	1.00	2	2	n/a	0	0	0.50	3	6	0.50	1	2
18	41.07	0.84	26	31	0.90	9	10	1.00	3	3	0.80	12	15	1.00	2	2	0.00	0	1	n/a	0	0	n/a	0	0
19	41.43	0.93	55	59	1.00	9	9	0.89	8	9	0.97	30	31	1.00	5	5	0.00	0	1	1.00	2	2	0.50	1	2
20	42.00	0.95	36	38	0.33	1	3	1.00	10	10	1.00	23	23	1.00	2	2	n/a	0	0	n/a	0	0	n/a	0	0
21	42.63	0.92	79	86	0.80	16	20	0.71	5	7	0.98	42	43	1.00	9	9	n/a	0	0	1.00	1	1	1.00	6	6
22	43.10	0.92	34	37	1.00	9	9	1.00	3	3	1.00	15	15	1.00	4	4	n/a	0	0	1.00	2	2	0.25	1	4
23	43.57	0.98	42	43	n/a	0	0	0.88	7	8	1.00	31	31	1.00	2	2	n/a	0	0	n/a	0	0	1.00	2	2
24	44.03	0.94	46	49	1.00	5	5	0.80	8	10	1.00	22	22	1.00	4	4	0.00	0	1	1.00	6	6	1.00	1	1
25	44.80	0.80	28	35	0.50	4	8	1.00	1	1	0.89	17	19	1.00	3	3	1.00	1	1	1.00	1	1	0.50	1	2

MZ1B	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	28.83	0.33	4	12	0.00	0	1	0.00	0	1	0.25	1	4	1.00	2	2	0.00	0	3	n/a	0	0	1.00	1	1
2	29.30	0.50	17	34	0.00	0	6	0.60	3	5	0.73	8	11	1.00	1	1	0.50	1	2	0.50	2	4	0.40	2	5
3	29.77	0.52	11	21	0.50	2	4	0.33	1	3	0.44	4	9	1.00	1	1	1.00	1	1	1.00	1	1	0.50	1	2
4°																									
5	31.90	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
6	32.37	0.56	39	70	0.57	8	14	0.44	4	9	0.67	24	36	n/a	0	0	0.33	1	3	n/a	0	0	0.25	2	8
7	32.83	0.53	24	45	0.36	4	11	0.25	1	4	0.62	13	21	0.00	0	1	1.00	1	1	1.00	2	2	0.60	3	5
8	33.30	0.76	65	86	0.93	40	43	0.00	0	7	0.83	20	24	0.00	0	1	n/a	0	0	1.00	1	1	0.40	4	10
9	33.87	0.49	23	47	0.18	2	11	0.80	4	5	0.57	12	21	n/a	0	0	0.00	0	1	1.00	1	1	0.50	4	8
10	34.57	0.64	27	42	0.73	11	15	0.14	1	7	0.75	9	12	1.00	1	1	n/a	0	0	0.50	1	2	0.80	4	5
11	35.03	0.40	17	42	0.25	4	16	0.20	2	10	0.78	7	9	n/a	0	0	0.67	2	3	1.00	1	1	0.33	1	3
12	37.80	0.75	12	16	n/a	0	0	0.33	1	3	0.86	6	7	1.00	2	2	1.00	1	1	1.00	1	1	0.50	1	2
13	38.27	1.00	16	16	1.00	3	3	1.00	2	2	1.00	10	10	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1
14	38.73	0.83	24	29	0.60	3	5	1.00	3	3	0.88	15	17	1.00	3	3	n/a	0	0	n/a	0	0	0.00	0	1
15	39.17	0.83	39	47	1.00	3	3	0.75	6	8	0.88	23	26	0.50	1	2	0.33	1	3	n/a	0	0	1.00	5	5
16	40.10	0.80	41	51	0.50	4	8	0.80	4	5	0.88	30	34	1.00	3	3	0.00	0	1	n/a	0	0	n/a	0	0
17	40.60	0.88	67	76	0.89	16	18	1.00	14	14	0.94	32	34	1.00	1	1	0.20	1	5	n/a	0	0	0.75	3	4
18	41.07	0.83	67	81	0.86	12	14	0.67	6	9	0.87	39	45	1.00	8	8	0.00	0	1	n/a	0	0	0.50	2	4
19	41.43	0.86	43	50	0.86	6	7	0.60	3	5	0.95	20	21	1.00	7	7	0.00	0	2	1.00	4	4	0.75	3	4
20	42.00	0.91	62	68	0.92	11	12	0.67	4	6	0.94	33	35	1.00	11	11	0.00	0	1	1.00	3	3	n/a	0	0
21	42.63	0.98	58	59	1.00	17	17	1.00	5	5	1.00	27	27	0.75	3	4	1.00	1	1	1.00	3	3	1.00	2	2
22	43.10	0.94	58	62	0.94	17	18	1.00	5	5	0.96	26	27	1.00	4	4	1.00	3	3	1.00	1	1	0.50	2	4
23	43.57	0.90	35	39	1.00	9	9	0.75	3	4	0.89	17	19	1.00	2	2	0.50	1	2	1.00	2	2	1.00	1	1
24	44.03	0.92	54	59	1.00	12	12	0.73	11	15	0.95	21	22	1.00	4	4	n/a	0	0	1.00	5	5	1.00	1	1
25	44.80	0.92	57	62	0.67	6	9	1.00	4	4	0.94	32	34	1.00	4	4	1.00	1	1	1.00	2	2	1.00	8	8

° The data for session 4 for MZ1B is missing.

MZ2A	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop %	cor	tot	do %	cor	tot	do in ques %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	24.77	0.88	7	8	1.00	1	1	1.00	1	1	0.80	4	5	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1
2	25.23	0.75	6	8	0.00	0	1	n/a	0	0	1.00	5	5	n/a	0	0	n/a	0	0	n/a	0	0	0.50	1	2
3	25.70	0.68	19	28	0.40	2	5	0.00	0	1	0.83	15	18	n/a	0	0	n/a	0	0	1.00	1	1	0.33	1	3
4	26.17	0.71	15	21	1.00	1	1	0.20	1	5	1.00	7	7	n/a	0	0	n/a	0	0	0.50	1	2	0.83	5	6
5	26.87	0.64	14	22	0.50	1	2	1.00	3	3	0.57	8	14	n/a	0	0	n/a	0	0	0.00	0	1	1.00	2	2
6	27.57	0.70	19	27	1.00	1	1	0.80	4	5	0.79	11	14	n/a	0	0	0.00	0	1	0.33	1	3	0.67	2	3
7	28.03	0.76	13	17	0.00	0	1	1.00	1	1	0.88	7	8	0.00	0	1	n/a	0	0	1.00	1	1	0.80	4	5
8	28.50	0.77	40	52	0.83	10	12	1.00	5	5	0.74	14	19	n/a	0	0	0.50	2	4	1.00	4	4	0.63	5	8
9	28.93	0.68	54	80	0.40	4	10	0.39	7	18	0.83	19	23	n/a	0	0	1.00	7	7	0.88	7	8	0.71	10	14
10	29.43	0.70	59	84	0.60	12	20	0.62	8	13	0.77	23	30	1.00	2	2	0.80	4	5	1.00	4	4	0.60	6	10
11	29.83	0.72	47	65	0.90	19	21	0.44	4	9	0.61	14	23	n/a	0	0	0.00	0	2	1.00	3	3	1.00	7	7
12	30.37	0.66	35	53	0.91	21	23	0.00	0	1	0.38	6	16	n/a	0	0	0.63	5	8	1.00	1	1	0.50	2	4
13	30.73	0.57	34	60	0.64	16	25	0.14	1	7	0.50	6	12	0.00	0	2	0.40	2	5	1.00	1	1	1.00	8	8
14	31.00	0.55	44	80	0.68	13	19	0.00	0	1	0.52	17	33	n/a	0	0	0.50	9	18	0.67	4	6	0.33	1	3
15	32.00	0.56	40	71	0.46	6	13	0.67	2	3	0.71	20	28	0.00	0	4	0.30	3	10	1.00	3	3	0.60	6	10
16	32.47	0.38	23	61	0.48	10	21	0.33	2	6	0.33	5	15	n/a	0	0	0.00	0	7	0.00	0	1	0.55	6	11
17	33.17	0.64	69	108	0.81	26	32	0.06	1	17	0.70	21	30	1.00	5	5	0.36	4	11	0.75	3	4	1.00	9	9
18	33.63	0.65	24	37	0.80	4	5	0.40	2	5	0.72	13	18	n/a	0	0	1.00	2	2	0.00	0	2	0.60	3	5
19	34.10	0.73	96	132	0.73	22	30	0.25	5	20	0.86	37	43	1.00	2	2	0.67	4	6	1.00	8	8	0.78	18	23
20	34.50	0.77	88	114	0.75	21	28	0.78	7	9	0.76	38	50	1.00	3	3	0.71	5	7	0.57	4	7	1.00	10	10
21	35.03	0.77	48	62	0.82	9	11	0.50	4	8	0.82	14	17	0.86	6	7	0.75	9	12	1.00	1	1	0.83	5	6
22	36.00	0.75	44	59	0.90	9	10	0.40	2	5	0.69	20	29	0.83	5	6	1.00	5	5	0.00	0	1	1.00	3	3
23	36.70	0.82	97	118	0.94	30	32	0.33	3	9	0.84	31	37	1.00	7	7	0.40	4	10	0.88	7	8	1.00	15	15
24	37.33	0.86	101	117	0.78	21	27	0.75	6	8	0.90	43	48	1.00	6	6	0.75	3	4	0.88	7	8	0.94	15	16
25	38.20	0.85	56	66	0.88	7	8	0.50	3	6	0.83	30	36	1.00	3	3	1.00	2	2	n/a	0	0	1.00	11	11
26	38.67	0.86	89	104	0.89	34	38	0.75	3	4	0.88	30	34	0.88	7	8	0.67	2	3	0.50	2	4	0.85	11	13
27	39.50	0.93	128	138	0.86	32	37	0.80	4	5	0.95	60	63	1.00	8	8	0.92	12	12	1.00	1	1	1.00	12	12
28	40.07	0.91	79	87	1.00	14	14	0.80	4	5	0.91	42	46	1.00	4	4	0.75	6	8	0.00	0	1	1.00	9	9
29	40.87	0.92	61	66	0.88	7	8	1.00	6	6	0.97	29	30	1.00	4	4	0.00	0	3	1.00	4	4	1.00	11	11
30	41.60	0.90	131	146	0.70	16	23	0.89	16	18	0.94	50	53	1.00	9	9	0.70	7	10	1.00	5	5	1.00	28	28

MZ2B	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	24.77	0.81	13	16	n/a	0	0	1.00	3	3	1.00	8	8	0.00	0	1	0.00	0	2	1.00	1	1	1.00	1	1
2	25.23	1.00	2	2	n/a	0	0	n/a	0	0	1.00	1	1	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1
3	25.70	1.00	4	4	1.00	1	1	n/a	0	0	1.00	2	2	n/a	0	0	n/a	0	0	1.00	1	1	n/a	0	0
4	26.17	0.62	10	16	0.00	0	3	1.00	1	1	1.00	7	7	1.00	1	1	1.00	1	1	0.00	0	1	0.00	0	2
5	26.87	0.50	1	2	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1	0.00	0	1
7	28.03	0.75	21	28	0.00	0	3	n/a	0	0	0.92	12	13	n/a	0	0	0.75	6	8	1.00	1	1	0.67	2	3
8	28.50	0.87	33	38	0.50	1	2	0.67	2	3	0.96	24	25	1.00	1	1	0.50	2	4	1.00	3	3	n/a	0	0
9	28.93	0.50	19	38	0.10	1	10	0.25	1	4	0.86	12	14	n/a	0	0	1.00	1	1	1.00	3	3	0.17	1	6
10	29.43	0.37	15	41	0.29	2	7	0.17	1	6	0.63	10	16	n/a	0	0	0.25	1	4	0.14	1	7	0.00	0	1
11	29.83	0.55	23	42	0.38	3	8	0.00	0	2	0.68	13	19	0.00	0	1	0.00	0	1	0.60	3	5	0.67	4	6
12	30.37	0.28	13	46	0.10	1	10	0.20	1	5	0.47	7	15	n/a	0	0	0.17	1	6	0.33	2	6	0.25	1	4
13	30.73	0.42	10	24	0.17	1	6	0.17	1	6	0.57	4	7	n/a	0	0	0.00	0	1	1.00	1	1	1.00	3	3
14	31.00	0.51	37	73	0.59	13	22	0.25	2	8	0.51	18	35	n/a	0	0	0.50	2	4	0.33	1	3	1.00	1	1
15	32.00	0.40	14	35	0.60	3	5	0.40	2	5	0.33	4	12	0.50	1	2	0.20	1	5	1.00	1	1	0.40	2	5
16	32.47	0.39	21	54	0.50	5	10	0.36	4	11	0.53	8	15	0.00	0	3	0.00	0	9	0.33	1	3	1.00	3	3
17	33.17	0.53	41	78	0.67	12	18	0.50	2	4	0.50	18	36	0.00	0	1	0.30	3	10	0.50	3	6	1.00	3	3
18	33.63	0.58	51	88	0.64	9	14	0.33	5	15	0.63	25	40	1.00	1	1	0.56	5	9	0.25	1	4	1.00	5	5
19	34.10	0.59	36	61	0.80	4	5	0.67	4	6	0.55	17	31	0.83	5	6	0.29	2	7	n/a	0	0	0.67	4	6
20	34.50	0.72	91	127	0.65	11	17	0.50	5	10	0.77	58	75	0.75	3	4	0.58	7	12	0.00	0	2	1.00	7	7
21	35.03	0.69	49	71	0.77	10	13	0.17	1	6	0.81	25	31	0.67	2	3	0.33	2	6	0.50	2	4	0.88	7	8
22	36.00	0.82	54	66	0.71	5	7	0.67	4	6	0.79	30	38	1.00	1	1	1.00	1	1	1.00	9	9	1.00	4	4
23	36.70	0.77	67	87	0.89	16	18	1.00	1	1	0.79	33	42	0.40	2	5	0.45	5	11	1.00	2	2	1.00	8	8
24	37.33	0.75	65	87	0.73	11	15	0.72	13	18	0.83	24	29	1.00	2	2	0.40	4	10	0.50	2	4	1.00	9	9
25	38.20	0.63	40	63	0.29	2	7	0.75	3	4	0.68	27	40	1.00	2	2	0.50	4	8	n/a	0	0	1.00	2	2
26	38.67	0.76	113	149	0.76	26	34	0.56	10	18	0.81	47	58	0.67	2	3	0.53	9	17	1.00	4	4	1.00	15	15
27	39.50	0.74	70	94	0.83	15	18	0.80	4	5	0.71	35	49	1.00	4	4	0.20	1	5	0.67	4	6	1.00	7	7
28	40.07	0.83	104	125	0.72	18	25	0.69	9	13	0.88	50	57	1.00	6	6	0.33	1	3	0.80	4	5	1.00	16	16
29	40.87	0.80	53	66	0.69	11	16	0.75	6	8	0.92	22	24	1.00	2	2	0.00	0	3	1.00	4	4	0.89	8	9
30	41.60	0.93	103	111	0.85	17	20	0.89	8	9	0.96	48	50	1.00	9	9	0.67	4	6	1.00	5	5	1.00	12	12

MZ3A	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	copbe %	cor	tot	do %	cor	tot	do in ques %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	22.37	0.50	1	2	n/a	0	0	n/a	0	0	1.00	1	1	n/a	0	0	n/a	0	0	n/a	0	0	0.00	0	1
2	22.83	9.99	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
5	24.20	9.99	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
6	24.90	0.28	5	18	0.20	1	5	0.33	1	3	0.40	2	5	n/a	0	0	n/a	0	0	n/a	0	0	0.20	1	5
7	25.43	0.50	8	16	0.80	4	5	1.00	1	1	0.50	2	4	n/a	0	0	0.00	0	1	n/a	0	0	0.20	1	5
8	26.60	0.32	6	19	0.17	1	6	n/a	0	0	0.43	3	7	n/a	0	0	1.00	1	1	n/a	0	0	0.20	1	5
9	27.30	0.36	4	11	0.00	0	1	n/a	0	0	0.40	2	5	n/a	0	0	1.00	1	1	n/a	0	0	0.25	1	4
10	27.77	0.30	12	40	0.16	3	19	0.10	1	10	0.67	4	6	n/a	0	0	1.00	2	2	1.00	2	2	0.00	0	1
11	28.67	0.18	7	38	0.00	0	8	0.17	2	12	0.44	4	9	0.00	0	2	n/a	0	0	0.00	0	1	0.17	1	6
12	29.40	0.76	38	50	0.00	0	1	0.88	7	8	0.77	30	39	n/a	0	0	n/a	0	0	n/a	0	0	0.50	1	2
13	29.87	0.79	26	33	n/a	0	0	1.00	3	3	0.77	23	30	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
14	30.57	0.71	27	38	0.14	1	7	0.75	9	12	0.88	15	17	n/a	0	0	1.00	1	1	1.00	1	1	n/a	0	0
15	31.03	0.73	41	56	0.40	2	5	0.62	13	21	0.87	26	30	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
16	31.50	0.70	57	81	0.33	4	12	0.57	8	14	0.87	40	46	0.67	2	3	n/a	0	0	1.00	2	2	0.25	1	4
17	32.43	0.54	13	24	0.00	0	4	0.38	3	8	1.00	9	9	0.50	1	2	n/a	0	0	n/a	0	0	0.00	0	1
18	32.90	0.69	37	54	0.10	1	10	0.40	4	10	0.95	19	20	0.50	1	2	n/a	0	0	1.00	11	11	1.00	1	1
19	33.37	0.74	20	27	n/a	0	0	0.90	9	10	0.67	8	12	1.00	1	1	n/a	0	0	0.67	2	3	0.00	0	1
20	33.83	0.31	15	48	0.04	1	23	0.00	0	6	0.75	12	16	n/a	0	0	n/a	0	0	1.00	2	2	0.00	0	1
21	34.50	0.73	24	33	0.43	3	7	0.75	3	4	0.85	11	13	1.00	3	3	n/a	0	0	1.00	2	2	0.50	2	4
22	36.10	0.76	38	50	1.00	1	1	0.58	11	19	0.94	16	17	1.00	7	7	n/a	0	0	1.00	1	1	0.40	2	5
23	36.53	0.59	17	29	0.43	3	7	0.00	0	3	0.56	5	9	1.00	3	3	0.00	0	1	1.00	3	3	1.00	3	3
24	37.00	0.77	49	64	0.63	5	8	0.59	10	17	0.87	26	30	1.00	3	3	n/a	0	0	0.67	2	3	1.00	3	3
25	37.93	0.74	28	38	0.43	3	7	0.75	6	8	0.91	10	11	1.00	2	2	0.75	3	4	0.50	1	2	0.75	3	4
26	39.33	0.70	26	37	0.00	0	1	0.57	8	14	0.78	14	18	1.00	1	1	n/a	0	0	1.00	1	1	1.00	2	2
27	39.87	0.84	43	51	1.00	2	2	0.80	20	25	0.92	12	13	1.00	4	4	0.67	2	3	1.00	1	1	0.67	2	3
28	40.97	0.73	11	15	0.50	1	2	0.75	3	4	0.75	3	4	n/a	0	0	0.67	2	3	1.00	1	1	1.00	1	1
29	41.63	0.89	51	57	0.00	0	1	0.90	9	10	0.94	32	34	1.00	2	2	1.00	2	2	1.00	3	3	0.60	3	5
30	42.87	0.83	44	53	0.73	8	11	0.71	5	7	0.85	22	26	1.00	2	2	1.00	2	2	1.00	2	2	1.00	3	3
31	43.80	0.94	68	72	0.85	11	13	0.93	13	14	0.97	31	32	1.00	5	5	1.00	2	2	1.00	1	1	1.00	5	5
32	45.27	0.90	71	79	0.83	5	6	0.76	16	21	0.97	35	36	0.75	3	4	1.00	2	2	1.00	2	2	1.00	8	8
33	46.33	0.86	43	50	0.67	8	12	0.80	4	5	0.91	21	23	1.00	3	3	n/a	0	0	1.00	1	1	1.00	6	6

MZ3B	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	22.37	0.67	2	3	n/a	0	0	n/a	0	0	0.50	1	2	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1
2	22.83	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
5	24.20	0.50	1	2	n/a	0	0	n/a	0	0	0.50	1	2	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
6	24.90	0.38	8	21	0.33	3	9	0.00	0	1	0.29	2	7	1.00	1	1	n/a	0	0	1.00	1	1	0.50	1	2
7	25.43	0.31	5	16	0.00	0	3	0.50	1	2	0.40	2	5	n/a	0	0	0.00	0	1	0.00	0	1	0.50	2	4
8	26.60	0.33	5	15	0.00	0	2	0.50	2	4	0.20	1	5	1.00	1	1	n/a	0	0	n/a	0	0	0.33	1	3
9	27.30	0.42	5	12	n/a	0	0	0.50	1	2	0.67	2	3	0.00	0	1	0.00	0	1	n/a	0	0	0.40	2	5
10	27.77	0.40	28	70	0.00	0	21	0.22	2	9	0.78	21	27	0.17	1	6	0.67	4	6	n/a	0	0	0.00	0	1
11	28.67	0.59	23	39	0.00	0	6	0.75	3	4	0.83	19	23	0.00	0	2	0.00	0	1	1.00	1	1	0.00	0	2
12	29.40	0.42	13	31	0.25	3	12	0.43	3	7	0.78	7	9	0.00	0	2	n/a	0	0	n/a	0	0	0.00	0	1
13	29.87	0.70	14	20	0.50	1	2	1.00	4	4	0.78	7	9	1.00	1	1	0.50	1	2	0.00	0	1	0.00	0	1
14	30.57	0.50	21	42	0.15	2	13	0.25	1	4	0.95	18	19	0.00	0	2	n/a	0	0	0.00	0	1	0.00	0	3
15	31.03	0.51	31	61	0.00	0	18	0.27	3	11	0.91	21	23	0.50	1	2	n/a	0	0	1.00	5	5	0.50	1	2
16	31.50	0.51	42	83	0.04	1	23	0.38	3	8	0.81	29	36	0.67	4	6	n/a	0	0	0.71	5	7	0.00	0	3
17	32.43	0.68	40	59	0.29	2	7	0.33	3	9	0.80	28	35	1.00	1	1	1.00	1	1	1.00	2	2	0.75	3	4
18	32.90	0.75	45	60	0.42	5	12	0.40	2	5	0.89	33	37	0.50	1	2	n/a	0	0	1.00	2	2	1.00	2	2
19	33.37	0.80	40	50	0.20	1	5	0.77	10	13	0.96	27	28	0.00	0	1	1.00	1	1	0.50	1	2	n/a	0	0
20	33.83	0.67	51	76	0.55	6	11	0.54	13	24	0.74	20	27	1.00	2	2	0.75	3	4	1.00	3	3	0.80	4	5
21	34.50	0.83	74	89	0.44	4	9	0.67	8	12	0.90	46	51	0.67	2	3	1.00	1	1	1.00	6	6	1.00	7	7
22	36.10	0.59	35	59	0.36	5	14	0.75	12	16	0.65	15	23	1.00	1	1	n/a	0	0	1.00	1	1	0.25	1	4
23	36.53	0.80	39	49	0.33	1	3	0.57	4	7	1.00	25	25	0.83	5	6	1.00	1	1	0.75	3	4	0.00	0	3
24	37.00	0.78	47	60	0.46	6	13	0.86	6	7	0.85	28	33	1.00	1	1	1.00	1	1	1.00	1	1	1.00	4	4
25	37.93	0.71	32	45	0.60	3	5	0.50	2	4	0.70	14	20	1.00	2	2	0.50	1	2	1.00	5	5	0.71	5	7
26	39.33	0.78	40	51	0.78	7	9	0.75	9	12	0.83	19	23	1.00	2	2	0.67	2	3	n/a	0	0	0.50	1	2
27	39.87	0.58	19	33	0.50	2	4	0.56	5	9	0.58	7	12	n/a	0	0	0.25	1	4	1.00	1	1	1.00	3	3
28	40.97	0.91	29	32	1.00	2	2	0.75	3	4	0.93	14	15	1.00	2	2	1.00	1	1	0.75	3	4	1.00	4	4
29	41.63	0.86	54	63	0.71	5	7	0.84	16	19	0.89	24	27	1.00	2	2	1.00	4	4	1.00	2	2	0.50	1	2
30	42.87	0.91	73	80	0.80	4	5	0.77	17	22	0.97	28	29	1.00	7	7	1.00	4	4	1.00	3	3	1.00	10	10
31	43.80	0.92	89	97	0.88	7	8	0.94	17	18	0.93	39	42	1.00	6	6	0.70	7	10	1.00	5	5	1.00	8	8
32	45.27	0.86	54	63	0.33	1	3	0.79	15	19	0.96	22	23	1.00	6	6	0.83	5	6	1.00	2	2	0.75	3	4
33	46.33	0.86	43	50	0.60	3	5	0.87	13	15	0.86	18	21	1.00	3	3	1.00	3	3	n/a	0	0	1.00	3	3

MZ4A	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
2	29.40	0.29	2	7	0.00	0	1	n/a	0	0	0.50	1	2	0.33	1	3	n/a	0	0	n/a	0	0	0.00	0	1
3	29.87	0.17	3	18	0.25	1	4	0.00	0	2	n/a	0	0	0.33	2	6	0.00	0	2	0.00	0	3	0.00	0	1
4	30.33	0.20	2	10	0.00	0	2	0.00	0	1	0.00	0	4	1.00	1	1	n/a	0	0	n/a	0	0	0.50	1	2
5	30.80	0.10	1	10	0.00	0	3	n/a	0	0	0.00	0	1	1.00	1	1	0.00	0	3	n/a	0	0	0.00	0	2
6	31.27	0.50	1	2	n/a	0	0	n/a	0	0	0.00	0	1	1.00	1	1	n/a	0	0	n/a	0	0	n/a	0	0
7	31.77	0.27	6	22	0.50	1	2	0.33	1	3	0.20	1	5	0.25	1	4	0.00	0	3	1.00	1	1	0.25	1	4
8	32.20	0.06	3	47	0.20	1	5	0.00	0	1	0.00	0	10	0.50	1	2	0.04	1	24	0.00	0	4	0.00	0	1
9	32.67	0.30	9	30	n/a	0	0	0.25	1	4	0.38	5	13	0.50	1	2	0.00	0	2	0.25	1	4	0.20	1	5
10	33.17	0.17	4	24	0.00	0	1	0.50	1	2	0.20	2	10	0.33	1	3	0.00	0	6	n/a	0	0	0.00	0	2
11	33.60	0.12	4	32	0.00	0	3	n/a	0	0	0.11	2	18	0.25	1	4	0.00	0	2	n/a	0	0	0.20	1	5
12	34.07	0.02	1	48	0.00	0	8	0.00	0	4	0.04	1	25	0.00	0	6	0.00	0	3	n/a	0	0	0.00	0	2
13	34.60	0.13	6	47	0.18	2	11	0.00	0	1	0.06	1	16	0.10	1	10	0.33	1	3	0.00	0	2	0.25	1	4
14	35.30	0.18	4	22	n/a	0	0	n/a	0	0	0.09	1	11	0.25	1	4	0.33	1	3	1.00	1	1	0.00	0	3
15	35.77	0.11	4	38	0.00	0	3	0.00	0	2	0.10	2	20	0.00	0	3	0.00	0	4	n/a	0	0	0.33	2	6
16	36.20	0.10	8	79	0.00	0	22	0.33	1	3	0.10	4	39	0.00	0	1	0.33	2	6	0.25	1	4	0.00	0	4
17	36.67	0.09	3	32	0.14	1	7	n/a	0	0	0.00	0	9	0.00	0	5	0.11	1	9	0.00	0	1	1.00	1	1
18	37.37	0.15	8	53	0.00	0	11	n/a	0	0	0.21	6	29	0.00	0	4	0.00	0	4	0.00	0	2	0.67	2	3
19	38.57	0.24	13	54	0.00	0	5	0.00	0	2	0.32	10	31	0.33	1	3	0.17	2	12	n/a	0	0	0.00	0	1
20	38.93	0.29	12	41	0.00	0	9	1.00	1	1	0.32	6	19	0.50	3	6	0.00	0	3	0.50	1	2	1.00	1	1
21	39.40	0.22	10	46	0.08	1	12	0.00	0	1	0.17	4	24	0.33	1	3	0.00	0	1	1.00	1	1	0.75	3	4
22	40.10	0.23	13	56	0.07	1	15	0.00	0	2	0.24	5	21	0.10	1	10	0.00	0	2	n/a	0	0	1.00	6	6
23	40.57	0.18	6	34	0.00	0	6	n/a	0	0	0.11	2	19	0.00	0	2	0.00	0	3	n/a	0	0	1.00	4	4
24	41.03	0.33	22	66	0.13	1	8	0.00	0	2	0.33	10	30	0.13	1	8	0.20	1	5	0.67	2	3	0.70	7	10
25	41.50	0.27	14	52	0.20	1	5	1.00	2	2	0.21	5	24	0.40	4	10	0.00	0	3	n/a	0	0	0.25	2	8
26	42.20	0.33	28	85	0.00	0	8	0.09	1	11	0.33	14	43	0.60	3	5	0.33	1	3	0.50	1	2	0.62	8	13
27	42.67	0.31	11	36	0.00	0	1	0.00	0	1	0.29	2	7	0.20	2	10	0.00	0	1	0.50	1	2	0.43	6	14
28	43.20	0.34	23	68	0.00	0	12	0.00	0	1	0.37	10	27	0.25	2	8	0.13	1	8	1.00	3	3	0.78	7	9
29	43.90	0.22	16	74	0.09	1	11	0.00	0	8	0.12	3	26	0.55	6	11	0.30	3	10	0.00	0	2	0.50	3	6
30	44.37	0.27	15	55	0.50	1	2	0.33	1	3	0.24	8	34	0.33	2	6	0.00	0	6	n/a	0	0	0.75	3	4
31	44.83	0.27	17	63	0.00	0	6	0.00	0	2	0.11	3	27	0.75	9	12	0.00	0	10	n/a	0	0	0.83	5	6
32	45.53	0.49	34	69	0.17	1	6	0.00	0	3	0.36	10	28	0.86	12	14	0.00	0	7	1.00	6	6	1.00	5	5
33	46.47	0.42	37	89	0.10	1	10	0.22	2	9	0.38	9	24	0.65	11	17	0.07	1	15	1.00	2	2	0.92	11	12

34	46.93	0.32	40	126	0.14	2	14	0.10	2	21	0.15	6	39	0.74	14	19	0.06	1	16	0.71	5	7	1.00	10	10
35	47.87	0.37	43	115	0.13	2	16	0.19	4	21	0.43	19	44	0.53	9	17	0.25	2	8	0.50	1	2	0.86	6	7
36	48.83	0.47	47	100	0.27	3	11	0.16	3	19	0.50	14	28	0.83	10	12	0.21	3	14	0.86	6	7	0.89	8	9
37	49.57	0.31	36	117	0.00	0	15	0.08	1	12	0.22	11	50	1.00	8	8	0.12	2	17	0.86	6	7	1.00	8	8
38	50.70	0.45	56	124	0.05	1	20	0.23	3	13	0.47	21	45	1.00	18	18	0.00	0	14	1.00	4	4	0.90	9	10
39	51.40	0.51	77	150	0.07	2	29	0.50	6	12	0.57	31	54	0.88	23	26	0.00	0	12	1.00	5	5	0.83	10	12
40	52.10	0.47	65	138	0.00	0	20	0.00	0	13	0.57	36	63	1.00	14	14	0.11	1	9	0.60	3	5	0.79	11	14
41	53.27	0.64	67	104	0.17	1	6	0.38	5	13	0.72	28	39	1.00	11	11	0.15	2	13	0.75	6	8	1.00	14	14
42	54.20	0.66	127	191	0.08	2	24	0.63	19	30	0.74	48	65	1.00	21	21	0.27	4	15	0.80	8	10	0.96	25	26
43	55.13	0.67	80	119	0.00	0	8	0.45	9	20	0.74	32	43	1.00	23	23	0.25	2	8	0.50	3	6	1.00	11	11
44	56.07	0.74	73	98	0.10	1	10	0.46	6	13	0.80	32	40	1.00	13	13	n/a	0	0	0.88	7	8	1.00	14	14

MZ4B	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
2	29.40	0.14	2	14	0.00	0	1	n/a	0	0	0.20	1	5	0.25	1	4	0.00	0	1	0.00	0	2	0.00	0	1
3	29.87	0.12	1	8	0.00	0	2	n/a	0	0	n/a	0	0	0.25	1	4	n/a	0	0	0.00	0	1	0.00	0	1
4	30.33	0.11	1	9	0.00	0	3	n/a	0	0	n/a	0	0	0.25	1	4	0.00	0	1	n/a	0	0	0.00	0	1
5	30.80	0.2	4	20	0.25	1	4	n/a	0	0	0.50	1	2	0.33	1	3	0.00	0	7	0.00	0	1	0.33	1	3
6	31.27	0.09	2	22	0.00	0	10	n/a	0	0	0.00	0	2	0.33	1	3	0.14	1	7	n/a	0	0	n/a	0	0
7	31.77	0.29	8	28	0.40	2	5	n/a	0	0	0.13	1	8	0.33	1	3	0.25	1	4	0.40	2	5	0.33	1	3
8	32.20	0.14	4	28	0.00	0	5	0.00	0	2	0.00	0	6	0.20	1	5	0.17	1	6	0.00	0	1	0.67	2	3
9	32.67	0.06	1	17	0.00	0	2	0.00	0	1	0.00	0	3	0.33	1	3	0.00	0	5	0.00	0	1	0.00	0	2
10	33.17	0.22	9	41	0.13	1	8	0.00	0	3	0.30	3	10	0.00	0	8	0.40	2	5	0.33	1	3	0.50	2	4
11	33.60	0.06	2	33	0.00	0	2	0.00	0	2	0.11	2	18	0.00	0	7	0.00	0	1	0.00	0	1	0.00	0	2
12	34.07	0.17	5	29	0.00	0	3	0.00	0	1	0.08	1	13	0.50	2	4	0.00	0	6	1.00	1	1	1.00	1	1
13	34.60	0.23	9	40	0.20	1	5	n/a	0	0	0.19	5	26	0.25	1	4	0.50	2	4	n/a	0	0	0.00	0	1
14	35.30	0.11	6	55	0.00	0	13	0.00	0	1	0.05	1	21	0.33	1	3	0.00	0	4	0.00	0	2	0.36	4	11
15	35.77	0.18	12	68	0.25	3	12	0.00	0	3	0.19	6	31	0.00	0	6	0.00	0	9	0.00	0	2	0.60	3	5
16	36.20	0.12	8	65	0.00	0	10	0.00	0	1	0.12	5	42	0.00	0	3	0.00	0	4	n/a	0	0	0.60	3	5
17	36.67	0.12	9	73	0.00	0	19	0.00	0	4	0.11	3	27	0.17	1	6	0.25	2	8	0.33	1	3	0.33	2	6
18	37.37	0.29	16	55	0.60	3	5	0.17	1	6	0.09	2	22	0.40	2	5	0.40	2	5	0.33	1	3	0.56	5	9
19	38.57	0.15	8	54	0.00	0	5	0.00	0	1	0.10	3	31	0.22	2	9	0.00	0	2	0.50	1	2	0.50	2	4
20	38.93	0.15	10	68	0.00	0	20	0.00	0	4	0.13	3	24	0.00	0	8	0.00	0	1	0.60	3	5	0.67	4	6
21	39.40	0.24	13	55	0.00	0	6	0.00	0	1	0.21	6	29	0.00	0	4	0.00	0	5	0.67	2	3	0.71	5	7
22	40.10	0.18	11	60	0.00	0	4	0.00	0	4	0.18	7	39	0.33	3	9	0.00	0	2	n/a	0	0	0.50	1	2
23	40.57	0.17	7	40	0.00	0	6	0.17	1	6	0.25	4	16	0.22	2	9	0.00	0	3	n/a	0	0	n/a	0	0
24	41.03	0.2	20	98	0.05	1	20	0.00	0	3	0.05	2	39	0.33	2	6	0.14	1	7	0.40	2	5	0.67	12	18
25	41.50	0.23	18	80	0.00	0	10	0.17	1	6	0.06	2	31	0.40	2	5	0.00	0	8	0.86	6	7	0.54	7	13
26	42.20	0.32	32	100	0.07	1	14	0.14	1	7	0.16	6	38	0.44	4	9	0.10	1	10	0.50	1	2	0.90	18	20
27	42.67	0.29	33	112	0.25	4	16	0.17	1	6	0.22	10	46	0.50	2	4	0.08	1	12	0.17	1	6	0.64	14	22
28	43.20	0.36	30	83	0.00	0	8	0.00	0	6	0.32	11	34	0.67	2	3	0.31	5	16	0.50	2	4	0.83	10	12
29	43.90	0.35	36	104	0.00	0	20	0.25	3	12	0.33	13	39	0.45	5	11	0.38	3	8	1.00	3	3	0.82	9	11
30	44.37	0.28	22	79	0.06	1	16	0.00	0	7	0.25	6	24	0.86	6	7	0.00	0	14	0.75	3	4	0.86	6	7
31	44.83	0.26	11	42	0.00	0	3	0.00	0	6	0.04	1	23	1.00	5	5	1.00	1	1	n/a	0	0	1.00	4	4
32	45.53	0.33	28	84	0.13	1	8	0.00	0	11	0.23	8	35	0.73	8	11	0.13	1	8	0.75	3	4	1.00	7	7
33	46.47	0.31	38	122	0.05	1	22	0.31	4	13	0.21	13	61	0.60	3	5	0.20	1	5	1.00	4	4	1.00	12	12

34	46.93	0.48	64	132	0.12	2	17	0.13	1	8	0.34	18	53	0.95	19	20	0.00	0	4	0.83	10	12	0.78	14	18
35	47.87	0.58	75	129	0.09	1	11	0.27	4	15	0.47	21	45	1.00	21	21	0.25	2	8	0.86	6	7	0.91	20	22
36	48.83	0.54	43	80	0.00	0	10	0.11	1	9	0.56	19	34	0.93	13	14	0.00	0	3	1.00	2	2	1.00	8	8
37	49.57	0.47	41	88	0.25	1	4	0.09	1	11	0.45	14	31	1.00	7	7	0.00	0	16	1.00	8	8	0.91	10	11
38	50.70	0.59	61	103	0.00	0	10	0.31	4	13	0.63	24	38	1.00	11	11	0.11	1	9	0.88	7	8	1.00	14	14
39	51.40	0.73	104	143	0.09	1	11	0.58	7	12	0.62	31	50	1.00	23	23	0.50	1	2	0.79	11	14	0.97	30	31
40	52.10	0.65	104	161	0.00	0	13	0.50	8	16	0.67	50	75	0.93	25	27	0.00	0	6	1.00	10	10	0.79	11	14
41	53.27	0.56	74	131	0.22	4	18	0.32	8	25	0.62	21	34	1.00	21	21	0.15	2	13	1.00	4	4	0.88	14	16
42	54.20	0.66	112	170	0.05	1	22	0.43	10	23	0.71	46	65	1.00	27	27	0.00	0	3	0.88	7	8	0.95	21	22
43	55.13	0.82	109	133	0.14	1	7	0.55	6	11	0.94	48	51	0.96	24	25	0.29	2	7	0.77	10	13	0.95	18	19
44	56.07	0.6	85	141	0.11	1	9	0.35	13	37	0.69	37	54	0.86	12	14	0.00	0	5	1.00	7	7	1.00	15	15

DZ1A	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	34.25	0.44	12	27	0.63	5	8	0.17	1	6	0.44	4	9	1.00	1	1	n/a	0	0	n/a	0	0	0.33	1	3
2	34.75	0.47	14	30	0.67	8	12	0.33	2	6	0.36	4	11	n/a	0	0	n/a	0	0	n/a	0	0	0.00	0	1
3	35.25	0.51	20	39	0.33	4	12	0.59	10	17	0.50	3	6	1.00	1	1	1.00	1	1	0.50	1	2	n/a	0	0
4	36.00	0.75	18	24	0.88	7	8	0.33	1	3	0.90	9	10	n/a	0	0	n/a	0	0	0.00	0	1	0.50	1	2
5	36.25	0.92	24	26	0.75	3	4	1.00	5	5	1.00	13	13	1.00	1	1	n/a	0	0	1.00	1	1	0.50	1	2
6	36.75	0.86	36	42	1.00	9	9	0.89	8	9	0.82	14	17	1.00	2	2	n/a	0	0	0.50	2	4	1.00	1	1
7	37.25	0.77	30	39	0.73	11	15	0.50	2	4	0.89	16	18	n/a	0	0	n/a	0	0	0.00	0	1	1.00	1	1
8	37.50	0.71	39	55	0.88	15	17	0.58	7	12	0.90	9	10	1.00	1	1	n/a	0	0	0.33	2	6	0.56	5	9
9	38.50	0.85	45	53	0.80	8	10	0.80	4	5	0.92	23	25	1.00	4	4	n/a	0	0	0.33	1	3	0.83	5	6
10	39.00	0.70	26	37	0.65	11	17	1.00	2	2	0.67	6	9	1.00	3	3	n/a	0	0	1.00	1	1	0.60	3	5
11	40.00	0.69	36	52	0.45	5	11	0.58	7	12	0.86	12	14	0.75	6	8	0.67	2	3	1.00	1	1	1.00	3	3
12	40.75	0.78	52	67	1.00	5	5	0.68	13	19	0.71	22	31	1.00	6	6	1.00	2	2	1.00	1	1	1.00	3	3
13	41.25	0.90	110	122	0.92	22	24	1.00	9	9	0.88	66	75	1.00	9	9	1.00	1	1	n/a	0	0	0.75	3	4
14	41.75	0.79	61	77	0.89	8	9	0.68	26	38	0.86	19	22	1.00	2	2	1.00	1	1	1.00	1	1	1.00	4	4
15	42.25	0.83	143	173	0.63	24	38	0.89	17	19	0.88	50	57	0.97	28	29	n/a	0	0	0.57	4	7	0.87	20	23
16	42.75	0.77	74	96	0.80	20	25	0.67	12	18	0.71	22	31	1.00	8	8	1.00	4	4	1.00	3	3	0.71	5	7
17	43.25	0.90	120	133	0.86	12	14	0.88	14	16	0.90	43	48	1.00	16	16	0.95	21	22	0.60	3	5	0.92	11	12
18	44.25	0.78	49	63	0.80	8	10	0.67	8	12	0.76	16	21	1.00	2	2	1.00	7	7	0.67	2	3	0.75	6	8
19	44.75	0.86	131	152	1.00	8	8	0.88	14	16	0.85	78	92	1.00	5	5	0.94	16	17	0.57	4	7	0.86	6	7
20	45.25	0.84	127	152	0.93	14	15	0.86	12	14	0.73	58	79	0.93	14	15	1.00	13	13	1.00	7	7	1.00	9	9
21	45.75	0.93	106	114	1.00	13	13	0.88	23	26	0.94	62	66	1.00	3	3	1.00	2	2	0.00	0	1	1.00	3	3
22	46.50	0.95	86	91	0.88	22	25	0.94	15	16	0.97	33	34	1.00	8	8	1.00	4	4	1.00	3	3	1.00	1	1
23	47.25	0.89	57	64	1.00	6	6	0.83	5	6	0.89	41	46	0.50	1	2	n/a	0	0	n/a	0	0	1.00	4	4
24	47.75	0.97	144	148	0.96	25	26	0.96	23	24	0.99	68	69	1.00	8	8	1.00	6	6	0.67	2	3	1.00	12	12

DZIB	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	34.25	0.25	1	4	0.00	0	1	n/a	0	0	0.50	1	2	n/a	0	0	n/a	0	0	n/a	0	0	0.00	0	1
2	34.75	0.71	15	21	0.75	3	4	1.00	1	1	0.64	7	11	n/a	0	0	1.00	3	3	n/a	0	0	0.50	1	2
3	35.25	0.75	3	4	1.00	1	1	0.00	0	1	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	1.00	2	2
4	36.00	0.43	3	7	1.00	1	1	n/a	0	0	0.33	2	6	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
5	36.25	0.77	10	13	0.00	0	2	1.00	2	2	1.00	7	7	n/a	0	0	n/a	0	0	n/a	0	0	0.50	1	2
6	36.75	0.38	3	8	0.00	0	1	0.50	2	4	0.00	0	2	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1
7	37.25	0.82	9	11	n/a	0	0	0.00	0	1	1.00	9	9	n/a	0	0	n/a	0	0	n/a	0	0	0.00	0	1
8	37.50	0.33	2	6	n/a	0	0	n/a	0	0	0.25	1	4	1.00	1	1	n/a	0	0	n/a	0	0	0.00	0	1
9	38.50	0.75	12	16	n/a	0	0	1.00	2	2	0.75	3	4	1.00	2	2	0.67	2	3	0.50	1	2	0.67	2	3
10	39.00	0.71	22	31	1.00	1	1	0.70	7	10	1.00	6	6	1.00	4	4	0.17	1	6	1.00	1	1	0.67	2	3
11	40.00	0.81	52	64	0.71	5	7	1.00	13	13	0.87	20	23	1.00	7	7	0.00	0	1	0.00	0	1	0.58	7	12
12	40.75	0.63	31	49	1.00	3	3	0.59	10	17	0.72	13	18	1.00	2	2	n/a	0	0	0.33	1	3	0.33	2	6
13	41.25	0.87	48	55	0.71	5	7	1.00	1	1	0.94	30	32	0.92	11	12	n/a	0	0	0.00	0	1	0.50	1	2
14	41.75	0.81	46	57	1.00	4	4	0.88	15	17	0.81	13	16	1.00	7	7	0.50	1	2	0.50	2	4	0.57	4	7
15	42.25	0.80	49	61	0.50	4	8	0.92	11	12	0.94	17	18	0.83	5	6	1.00	2	2	0.50	1	2	0.69	9	13
16	42.75	0.89	65	73	0.67	4	6	0.90	9	10	1.00	30	30	1.00	9	9	0.83	5	6	0.50	2	4	0.75	6	8
17	43.25	0.87	86	99	0.85	11	13	0.90	9	10	0.88	45	51	1.00	11	11	0.83	5	6	0.50	1	2	0.67	4	6
18	44.25	0.74	55	74	0.33	4	12	0.71	10	14	0.88	22	25	1.00	7	7	0.50	2	4	0.50	1	2	0.90	9	10
19	44.75	0.73	11	15	0.67	2	3	1.00	1	1	0.71	5	7	1.00	1	1	n/a	0	0	1.00	1	1	0.50	1	2
20	45.25	0.83	59	71	1.00	7	7	0.73	11	15	0.87	33	38	1.00	1	1	n/a	0	0	0.75	3	4	0.67	4	6
21	45.75	0.95	82	86	1.00	5	5	1.00	28	28	0.97	36	37	1.00	3	3	1.00	1	1	1.00	3	3	0.67	6	9
22	46.50	0.89	63	71	0.85	11	13	1.00	16	16	0.96	22	23	1.00	5	5	0.75	3	4	0.83	5	6	0.25	1	4
23	47.25	0.93	69	74	1.00	3	3	1.00	8	8	0.93	28	30	1.00	5	5	0.90	9	10	0.92	12	13	0.80	4	5
24	47.75	0.92	54	59	1.00	4	4	1.00	11	11	1.00	24	24	1.00	1	1	1.00	1	1	0.86	6	7	0.64	7	11

DZ2A	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
6*	28.23	0.45	5	11	1.00	1	1	0.67	2	3	0.67	2	3	n/a	0	0	n/a	0	0	0.00	0	1	0.00	0	3
7	29.03	0.90	19	21	1.00	7	7	0.50	1	2	1.00	8	8	n/a	0	0	1.00	3	3	n/a	0	0	0.00	0	1
8	29.73	0.89	25	28	1.00	2	2	n/a	0	0	0.89	17	19	1.00	1	1	1.00	5	5	n/a	0	0	0.00	0	1
9	30.20	0.56	25	45	0.00	0	5	1.00	5	5	0.67	8	12	1.00	2	2	0.80	4	5	0.00	0	2	0.43	6	14
10	31.13	0.74	31	42	0.20	1	5	0.75	6	8	0.89	17	19	n/a	0	0	0.75	6	8	n/a	0	0	0.50	1	2
11	31.60	0.58	19	33	0.14	1	7	n/a	0	0	0.71	12	17	n/a	0	0	0.75	3	4	0.33	1	3	1.00	2	2
12	32.07	0.67	54	81	0.12	2	17	0.78	7	9	0.88	37	42	0.33	1	3	1.00	3	3	0.00	0	1	0.67	4	6
13	32.43	0.71	60	84	0.60	9	15	0.75	12	16	0.77	23	30	1.00	1	1	0.64	7	11	0.33	1	3	0.88	7	8
14	32.90	0.72	49	68	0.40	6	15	0.77	10	13	0.83	19	23	1.00	2	2	0.60	3	5	1.00	5	5	0.80	4	5
15	33.37	0.58	54	93	0.06	1	16	0.72	13	18	0.76	28	37	1.00	1	1	0.50	3	6	0.40	2	5	0.60	6	10
16	34.07	0.61	35	57	0.20	2	10	0.67	4	6	0.74	14	19	1.00	2	2	0.29	2	7	1.00	3	3	0.80	8	10
17	34.77	0.76	95	125	0.33	8	24	0.91	30	33	0.90	36	40	1.00	1	1	0.71	5	7	1.00	3	3	0.71	12	17
18	35.47	0.72	84	117	0.65	24	37	0.68	23	34	0.88	22	25	n/a	0	0	0.33	1	3	0.80	8	10	0.75	6	8
19	36.40	0.81	34	42	0.56	5	9	0.80	4	5	0.87	13	15	1.00	1	1	1.00	2	2	1.00	3	3	0.86	6	7
20	37.57	0.83	50	60	0.56	5	9	0.88	7	8	0.96	27	28	1.00	2	2	1.00	3	3	0.67	2	3	0.57	4	7
21	38.50	0.72	69	96	0.44	8	18	0.94	16	17	0.92	24	26	1.00	10	10	n/a	0	0	0.50	2	4	0.43	9	21
22	39.80	0.77	46	60	0.67	2	3	0.63	5	8	0.82	23	28	1.00	8	8	0.00	0	1	1.00	5	5	0.43	3	7
23	40.60	0.76	60	79	0.50	4	8	0.76	13	17	0.92	24	26	1.00	7	7	0.00	0	1	0.00	0	3	0.71	12	17
24	41.77	0.79	44	56	0.88	7	8	0.75	12	16	0.78	14	18	1.00	6	6	1.00	1	1	n/a	0	0	0.57	4	7
25	43.40	0.76	42	55	0.73	11	15	0.60	3	5	1.00	19	19	1.00	2	2	1.00	1	1	0.20	1	5	0.63	5	8
26	44.33	0.84	97	116	0.82	14	17	0.86	24	28	1.00	32	32	1.00	5	5	0.67	2	3	0.70	7	10	0.62	13	21
27	45.53	0.93	54	58	1.00	10	10	0.83	5	6	0.95	19	20	1.00	3	3	1.00	2	2	0.86	6	7	0.90	9	10

* The first 5 sessions were not used for the DZ2 pair because the children had very few lines.

DZ2B	Age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
6	28.23	0.43	15	35	0.40	6	15	1.00	1	1	0.47	7	15	1.00	1	1	0.00	0	1	n/a	0	0	0.00	0	2
7	29.03	0.92	11	12	0.50	1	2	1.00	1	1	1.00	7	7	1.00	1	1	1.00	1	1	n/a	0	0	n/a	0	0
8	29.73	0.37	10	27	0.25	3	12	0.80	4	5	0.50	2	4	n/a	0	0	0.00	0	1	0.00	0	2	0.33	1	3
9	30.20	0.68	42	62	0.20	1	5	0.67	6	9	0.89	33	37	0.00	0	1	n/a	0	0	0.00	0	1	0.22	2	9
10	31.13	0.71	5	7	0.00	0	1	n/a	0	0	1.00	3	3	0.50	1	2	n/a	0	0	n/a	0	0	1.00	1	1
11	31.60	0.43	13	30	0.00	0	4	n/a	0	0	0.67	10	15	0.00	0	4	0.33	1	3	0.00	0	2	1.00	2	2
12	32.07	0.61	39	64	0.31	4	13	0.67	2	3	0.88	22	25	0.67	2	3	0.80	4	5	0.50	2	4	0.27	3	11
13	32.43	0.78	77	99	0.30	3	10	0.57	4	7	0.96	55	57	0.29	2	7	0.67	8	12	1.00	1	1	0.80	4	5
14	32.90	0.74	66	89	0.42	5	12	0.59	10	17	0.95	36	38	0.67	4	6	0.80	4	5	1.00	3	3	0.50	4	8
15	33.37	0.53	42	79	0.00	0	10	0.25	3	12	0.83	25	30	0.33	1	3	0.17	1	6	0.78	7	9	0.56	5	9
16	34.07	0.51	35	68	0.05	1	19	0.67	2	3	0.82	14	17	0.86	6	7	0.67	8	12	0.00	0	1	0.44	4	9
17	34.77	0.72	64	89	0.07	1	15	0.73	8	11	0.89	39	44	0.00	0	1	1.00	4	4	0.50	1	2	0.92	11	12
18	35.47	0.70	55	79	0.13	2	15	0.82	9	11	0.97	34	35	0.00	0	1	0.50	1	2	0.67	2	3	0.58	7	12
19	36.40	0.75	42	56	0.60	3	5	0.89	8	9	0.82	23	28	1.00	2	2	0.33	1	3	0.75	3	4	0.40	2	5
20	37.57	0.79	61	77	0.80	8	10	0.75	3	4	0.93	26	28	0.67	6	9	1.00	2	2	0.80	4	5	0.63	12	19
21	38.50	0.79	52	66	0.87	13	15	0.90	9	10	0.90	18	20	0.57	4	7	1.00	3	3	0.75	3	4	0.29	2	7
22	39.80	0.84	82	98	0.87	13	15	0.81	13	16	0.91	29	32	1.00	8	8	0.40	2	5	1.00	7	7	0.67	10	15
23	40.60	0.94	61	65	1.00	5	5	1.00	3	3	0.95	35	37	1.00	10	10	0.60	3	5	1.00	1	1	1.00	4	4
24	41.77	0.98	65	66	1.00	6	6	1.00	8	8	1.00	20	20	1.00	15	15	1.00	4	4	1.00	2	2	0.91	10	11
25	43.40	0.86	25	29	1.00	2	2	1.00	2	2	1.00	12	12	1.00	3	3	n/a	0	0	0.50	1	2	0.63	5	8
26	44.33	0.94	63	67	1.00	11	11	1.00	8	8	0.93	25	27	1.00	6	6	0.67	2	3	1.00	6	6	0.83	5	6
27	45.53	0.84	11	131	0.94	16	17	0.90	19	21	0.90	46	51	1.00	11	11	0.50	8	16	0.67	6	9	0.67	4	6
28	47.10	0.92	48	52	0.92	11	12	1.00	6	6	0.89	17	19	1.00	5	5	0.86	6	7	1.00	1	1	1.00	2	2

DZ3A	age in months	averag e %	correct	total	3s %	co r	tot	auxbe %	co r	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	26.83	1.00	4	4	n/a	0	0	n/a	0	0	1.00	4	4	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
2	27.30	0.25	1	4	0.00	0	1	n/a	0	0	n/a	0	0	n/a	0	0	0.33	1	3	n/a	0	0	n/a	0	0
3	27.77	9.99	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
4	28.23	0.67	2	3	0.00	0	1	n/a	0	0	1.00	1	1	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1
5	28.70	0.00	0	9	0.00	0	4	n/a	0	0	0.00	0	3	n/a	0	0	0.00	0	2	n/a	0	0	n/a	0	0
6	29.17	0.18	2	11	0.00	0	2	0	0	2	0.00	0	5	n/a	0	0	n/a	0	0	n/a	0	0	1.00	2	2
7	29.63	0.00	0	9	0.00	0	6	n/a	0	0	0.00	0	2	0.00	0	1	n/a	0	0	n/a	0	0	n/a	0	0
8	30.07	0.44	7	16	0.33	1	3	n/a	0	0	0.60	3	5	0.00	0	1	0.00	0	1	0.50	2	4	0.50	1	2
9	30.80	0.14	1	7	n/a	0	0	n/a	0	0	0.00	0	2	n/a	0	0	0.00	0	4	n/a	0	0	1.00	1	1
10	31.27	0.05	1	19	0.00	0	3	n/a	0	0	0.00	0	6	0.00	0	6	0.00	0	3	n/a	0	0	1.00	1	1
11	32.20	0.07	1	14	0.00	0	1	n/a	0	0	0.09	1	11	0.00	0	2	n/a	0	0	n/a	0	0	n/a	0	0
12	32.67	0.15	2	13	0.33	1	3	n/a	0	0	0.14	1	7	0.00	0	2	n/a	0	0	n/a	0	0	0.00	0	1
13	33.67	0.19	6	31	0.14	1	7	0	0	1	0.07	1	14	0.50	1	2	0.00	0	2	1.00	1	1	0.50	2	4
14	34.53	0.35	11	31	0.50	2	4	0	0	1	0.21	3	14	0.50	2	4	0.00	0	1	0.67	2	3	0.50	2	4
15	35.00	0.31	8	26	0.00	0	3	0	0	1	0.36	4	11	0.50	2	4	n/a	0	0	0.00	0	2	0.40	2	5
16	35.70	0.40	10	25	0.33	1	3	0	0	1	0.29	4	14	0.50	1	2	0.00	0	1	n/a	0	0	1.00	4	4
17	35.93	0.25	2	8	0.00	0	2	0	0	2	0.33	1	3	n/a	0	0	n/a	0	0	n/a	0	0	1.00	1	1
18	36.63	0.44	15	34	0.20	1	5	n/a	0	0	0.47	7	15	0.50	1	2	0.33	1	3	n/a	0	0	0.56	5	9
19	37.33	0.43	16	37	0.00	0	5	0.5	1	2	0.33	6	18	0.00	0	2	0.86	6	7	n/a	0	0	1.00	3	3
20	38.03	0.41	9	22	0.00	0	3	0	0	4	0.50	5	10	1.00	1	1	0.50	1	2	n/a	0	0	1.00	2	2
21	38.50	0.47	19	40	0.25	1	4	0	0	3	0.30	6	20	1.00	2	2	n/a	0	0	1.00	2	2	0.89	8	9
22	38.97	0.56	9	16	1.00	1	1	n/a	0	0	0.38	3	8	n/a	0	0	n/a	0	0	0.00	0	1	0.83	5	6
23	39.67	0.60	9	15	0.50	1	2	1	1	1	0.38	3	8	1.00	2	2	n/a	0	0	1.00	1	1	1.00	1	1
24	40.60	0.31	13	42	0.00	0	1	0	0	8	0.28	7	25	n/a	0	0	0.50	1	2	0.75	3	4	1.00	2	2
25	40.83	0.43	26	60	0.33	1	3	0.25	2	8	0.37	7	19	0.33	1	3	0.50	1	2	0.33	1	3	0.59	13	22
26	41.30	0.68	32	47	0.50	1	2	0	0	2	0.33	2	6	1.00	9	9	0.25	1	4	1.00	6	6	0.72	13	18
27	41.83	0.38	14	37	0.00	0	2	0.18	2	11	0.33	3	9	1.00	4	4	0.00	0	1	0.50	1	2	0.50	4	8
28	42.30	0.51	18	35	0.00	0	1	0.29	2	7	0.3	3	10	1.00	5	5	0.00	0	1	0.40	2	5	1.00	6	6
29	42.77	0.30	12	40	n/a	0	0	0	0	10	0.24	4	17	0.67	2	3	0.00	0	2	1.00	2	2	0.67	4	6
30	43.23	0.45	27	60	0.20	1	5	0	0	14	0.5	10	20	1.00	2	2	n/a	0	0	0.00	0	1	0.78	14	18
31	43.70	0.49	17	35	n/a	0	0	0	0	1	0.32	8	25	1.00	7	7	n/a	0	0	n/a	0	0	1.00	2	2
32	44.40	0.35	16	46	0.00	0	2	0	0	16	0.53	10	19	1.00	5	5	0.00	0	3	n/a	0	0	1.00	1	1

33	44.87	0.33	29	89	0.25	3	12	0.10	2	21	0.29	8	28	1.00	5	5	0.00	0	1	0.40	2	5	0.53	9	17
34	46.03	0.50	20	40	0.50	1	2	0	0	3	0.43	10	23	1.00	3	3	0.50	1	2	1.00	1	1	0.67	4	6
35	46.50	0.50	21	42	n/a	0	0	0	0	4	0.36	4	11	1.00	1	1	0.00	0	1	0.64	7	11	0.64	9	14
36	46.97	0.51	21	41	0.00	0	1	0.08	1	12	0.25	2	8	1.00	4	4	n/a	0	0	1.00	2	2	0.86	12	14
37	47.43	0.54	19	35	0.00	0	1	0.3	3	10	0.43	3	7	1.00	6	6	1.00	1	1	0.75	3	4	0.50	3	6
38	47.90	0.33	19	58	0.17	1	6	0.05	1	20	0.36	8	22	1.00	4	4	n/a	0	0	1.00	3	3	0.67	2	3
39	48.60	0.71	92	130	0.38	3	8	0.55	6	11	0.71	41	58	1.00	10	10	0.00	0	1	0.94	16	17	0.64	16	25
40	49.30	0.59	19	32	0.67	2	3	0	0	8	0.69	9	13	1.00	2	2	1.00	1	1	1.00	1	1	1.00	4	4
41	50.00	0.84	51	61	0.40	2	5	0.71	5	7	0.82	14	17	1.00	4	4	n/a	0	0	1.00	7	7	0.90	19	21
42	50.70	0.61	28	46	0.67	4	6	0.2	2	10	0.56	9	16	1.00	1	1	1.00	2	2	0.67	2	3	1.00	8	8

DZ3B	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	26.83	9.99	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
2	27.30	9.99	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
3	27.77	0	0	2	n/a	0	0	n/a	0	0	0.00	0	1	n/a	0	0	n/a	0	0	0.00	0	1	n/a	0	0
4	28.23	0.07	1	14	0.00	0	7	0.00	0	1	0.00	0	2	n/a	0	0	0.00	0	1	0.50	1	2	0.00	0	1
5	28.70	0	0	3	n/a	0	0	n/a	0	0	0.00	0	1	n/a	0	0	0.00	0	1	0.00	0	1	n/a	0	0
6	29.17	0	0	1	0.00	0	1	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
7	29.63	0.14	2	14	0.13	1	8	n/a	0	0	0.00	0	1	n/a	0	0	0.00	0	3	n/a	0	0	0.50	1	2
8	30.07	0	0	10	n/a	0	0	n/a	0	0	0.00	0	1	0.00	0	3	0.00	0	1	0.00	0	4	0.00	0	1
9	30.80	0	0	13	0.00	0	6	n/a	0	0	0.00	0	2	n/a	0	0	0.00	0	1	0.00	0	1	0.00	0	3
10	31.27	0.09	3	33	0.00	0	11	n/a	0	0	0.07	1	14	0.33	1	3	0.00	0	1	0.50	1	2	0.00	0	2
11	32.20	0.05	1	21	0.00	0	8	n/a	0	0	0.00	0	5	0.00	0	1	0.33	1	3	n/a	0	0	0.00	0	4
12	32.67	0	0	9	0.00	0	3	n/a	0	0	0.00	0	2	0.00	0	1	0.00	0	3	n/a	0	0	n/a	0	0
13	33.67	0.14	4	28	0.00	0	6	0.00	0	1	0.08	1	12	n/a	0	0	0.00	0	1	0.50	1	2	0.33	2	6
14	34.53	0.08	2	24	0.00	0	2	0.00	0	1	0.09	1	11	0.00	0	1	0.00	0	2	0.00	0	1	0.17	1	6
15	35.00	0.07	1	14	0.00	0	7	n/a	0	0	0.33	1	3	0.00	0	1	0.00	0	1	0.00	0	1	0.00	0	1
16	35.70	0	0	2	0.00	0	1	0.00	0	1	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0	n/a	0	0
17	35.93	0.33	4	12	0.00	0	2	n/a	0	0	0.29	2	7	0.00	0	1	n/a	0	0	n/a	0	0	1.00	2	2
18	36.63	0.32	16	50	n/a	0	0	n/a	0	0	0.15	2	13	0.00	0	4	0.00	0	1	0.38	3	8	0.46	11	24
19	37.33	0.3	8	27	0.00	0	3	n/a	0	0	0.22	2	9	0.33	2	6	0.00	0	1	0.00	0	2	0.67	4	6
20	38.03	0.13	5	39	0.00	0	12	0.00	0	3	0.13	2	16	0.50	2	4	0.00	0	3	n/a	0	0	1.00	1	1
21	38.50	0.19	5	26	0.00	0	5	0.00	0	8	0.44	4	9	0.00	0	1	n/a	0	0	n/a	0	0	0.33	1	3
22	38.97	0.2	2	10	0.00	0	1	n/a	0	0	0.20	1	5	n/a	0	0	0.00	0	1	n/a	0	0	0.33	1	3
23	39.67	0.22	8	37	0.20	1	5	0.18	2	11	0.21	3	14	1.00	1	1	0.00	0	3	0.00	0	2	1.00	1	1
24	40.60	0.45	9	20	n/a	0	0	0.00	0	4	0.67	2	3	0.40	2	5	0.50	1	2	0.00	0	1	0.80	4	5
25	40.83	0.41	17	41	0.33	1	3	0.25	1	4	0.17	1	6	0.78	7	9	n/a	0	0	1.00	1	1	0.33	6	18
26	41.30	0.31	11	36	0.25	1	4	0.11	1	9	0.29	2	7	0.50	4	8	0.00	0	2	0.33	1	3	0.67	2	3
27	41.83	0.32	17	53	0.00	0	2	0.00	0	1	0.29	5	17	1.00	6	6	n/a	0	0	0.38	3	8	0.16	3	19
28	42.30	0.37	11	30	0.00	0	3	0.00	0	5	0.25	2	8	1.00	4	4	n/a	0	0	0.00	0	3	0.71	5	7
29	42.77	0.47	14	30	0.40	2	5	0.00	0	1	0.33	5	15	1.00	2	2	n/a	0	0	0.75	3	4	0.67	2	3
30	43.23	0.33	14	42	0.00	0	2	0.20	1	5	0.20	4	20	1.00	3	3	0.00	0	1	n/a	0	0	0.55	6	11
31	43.70	0.33	15	46	0.00	0	2	0.13	1	8	0.19	3	16	1.00	5	5	n/a	0	0	0.00	0	4	0.55	6	11
32	44.40	0.4	4	10	0.00	0	1	0.00	0	1	0.00	0	1	1.00	2	2	n/a	0	0	n/a	0	0	0.40	2	5

33	44.87	0.41	29	71	0.10	1	10	0.00	0	16	0.33	4	12	0.95	18	19	n/a	0	0	0.60	3	5	0.33	3	9
34	46.03	0.44	31	71	0.19	3	16	0.14	1	7	0.29	7	24	1.00	11	11	0.50	2	4	0.67	2	3	0.83	5	6
35	46.50	0.44	14	32	0.33	2	6	0.00	0	3	0.25	3	12	1.00	2	2	n/a	0	0	0.00	0	1	0.88	7	8
36	46.97	0.54	34	63	0.00	0	5	0.00	0	2	0.18	2	11	1.00	7	7	n/a	0	0	0.55	6	11	0.70	19	27
37	47.43	0.56	9	16	1.00	1	1	0.00	0	2	0.00	0	3	1.00	5	5	0.33	1	3	n/a	0	0	1.00	2	2
38	47.90	0.29	10	35	0.33	1	3	0.10	1	10	0.17	2	12	1.00	2	2	n/a	0	0	n/a	0	0	0.50	4	8
39	48.60	0.49	21	43	0.11	1	9	0.18	2	11	0.29	2	7	1.00	11	11	n/a	0	0	1.00	1	1	1.00	4	4
40	49.30	0.35	16	46	0.00	0	10	0.18	2	11	0.53	8	15	1.00	4	4	0.00	0	2	n/a	0	0	0.50	2	4
41	50.00	0.45	25	55	0.00	0	12	0.57	4	7	0.23	3	13	1.00	4	4	n/a	0	0	0.60	3	5	0.79	11	14
42	50.70	0.4	22	55	0.11	1	9	0.40	8	20	0.33	3	9	1.00	6	6	0.00	0	1	0.00	0	1	0.44	4	9

DZ4A	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	32.57	0.80	39	49	n/a	0	0	0.57	4	7	0.83	29	35	1.00	1	1	0.83	5	6	n/a	0	0	n/a	0	0
2	33.03	0.93	55	59	0.83	5	6	0.94	15	16	0.96	26	27	1.00	4	4	1.00	2	2	n/a	0	0	0.75	3	4
3	33.27	0.73	22	30	0.00	0	3	0.67	2	3	0.79	11	14	1.00	1	1	[illegible]	4	5	1.00	1	1	1.00	3	3
4	33.73	0.84	58	69	0.83	5	6	0.82	9	11	0.86	30	35	1.00	6	6	0.70	7	10	n/a	0	0	1.00	1	1
5	34.20	0.88	67	76	0.50	4	8	0.87	13	15	0.95	36	38	1.00	2	2	1.00	4	4	n/a	0	0	0.89	8	9
6	34.80	0.85	127	150	0.50	13	26	0.93	13	14	0.96	81	84	1.00	6	6	0.65	11	17	1.00	1	1	1.00	2	2
7	35.17	0.82	108	131	0.44	11	25	0.88	7	8	0.93	79	85	1.00	5	5	0.67	2	3	0.00	0	1	1.00	4	4
8	35.63	0.92	71	77	1.00	10	10	0.86	6	7	0.97	37	38	1.00	6	6	1.00	2	2	n/a	0	0	0.71	10	14
9	36.10	0.91	170	187	0.86	12	14	0.89	8	9	0.95	122	129	1.00	15	15	0.67	4	6	0.50	1	2	0.67	8	12
10	37.03	0.80	109	137	0.69	11	16	0.68	13	19	0.90	60	67	1.00	3	3	0.86	6	7	1.00	6	6	0.53	10	19
11	39.20	0.98	124	127	0.89	8	9	0.95	19	20	1.00	67	67	1.00	16	16	1.00	8	8	0.50	1	2	1.00	5	5
12	39.67	0.96	128	134	0.88	7	8	0.95	20	21	1.00	54	54	1.00	21	21	1.00	10	10	0.00	0	1	0.84	16	19
14*	40.60	0.96	64	67	0.70	7	10	1.00	7	7	1.00	38	38	1.00	3	3	1.00	7	7	n/a	0	0	1.00	2	2
15	41.47	0.96	152	159	0.96	22	23	0.94	15	16	0.99	77	78	1.00	16	16	0.82	14	17	0.00	0	1	1.00	8	8
16	42.00	0.97	145	149	1.00	21	21	0.90	19	21	0.99	68	69	1.00	10	10	1.00	22	22	1.00	3	3	0.67	2	3
17	42.47	1.00	28	28	1.00	5	5	1.00	1	1	1.00	19	19	1.00	1	1	1.00	2	2	n/a	0	0	n/a	0	0

* session 13 is missing for the DZ4 pair because no taping was done during this visit.

DZ4B	age in months	average %	correct	total	3s %	cor	tot	auxbe %	cor	tot	cop be %	cor	tot	do %	cor	tot	do in ques. %	cor	tot	ed %	cor	tot	ptirr %	cor	tot
1	32.57	0.54	43	80	0.67	2	3	0.33	2	6	0.63	30	48	0.80	4	5	0.00	0	2	0.67	2	3	0.23	3	13
2	33.03	0.70	44	63	0.33	1	3	0.69	9	13	0.82	27	33	1.00	1	1	0.33	1	3	1.00	3	3	0.29	2	7
3	33.27	0.73	40	55	n/a	0	0	0.53	10	19	0.82	23	28	1.00	4	4	n/a	0	0	n/a	0	0	0.75	3	4
4	33.73	0.74	39	53	1.00	2	2	0.57	8	14	0.81	26	32	1.00	1	1	0.50	1	2	0.00	0	1	1.00	1	1
5	34.20	0.83	55	66	0.50	3	6	0.83	5	6	0.87	40	46	0.67	2	3	1.00	3	3	n/a	0	0	1.00	2	2
6	34.80	0.89	49	55	1.00	8	8	0.80	8	10	0.92	23	25	0.75	3	4	1.00	4	4	n/a	0	0	0.75	3	4
7	35.17	0.84	71	85	0.75	12	16	0.60	3	5	0.91	49	54	1.00	2	2	0.60	3	5	0.00	0	1	1.00	2	2
8	35.63	0.73	60	82	0.46	6	13	0.69	9	13	0.94	31	33	1.00	2	2	1.00	2	2	0.00	0	2	0.59	10	17
9	36.10	0.89	97	109	0.86	6	7	0.83	10	12	0.96	71	74	n/a	0	0	0.83	5	6	0.00	0	2	0.63	5	8
10	37.03	0.93	110	118	0.92	11	12	0.83	10	12	0.97	76	78	1.00	4	4	1.00	6	6	0.50	1	2	0.50	2	4
11	39.20	0.96	151	157	0.85	11	13	1.00	12	12	0.99	91	92	1.00	8	8	0.93	13	14	1.00	2	2	0.88	14	16
12	39.67	0.96	110	114	1.00	6	6	0.95	18	19	1.00	60	60	1.00	4	4	1.00	10	10	0.00	0	1	0.86	12	14
14*																									
15	41.47	0.99	109	110	0.90	9	10	1.00	12	12	1.00	54	54	1.00	7	7	1.00	23	23	1.00	1	:	1.00	3	3
16	42.00	0.92	84	91	0.70	7	10	1.00	15	15	1.00	33	33	1.00	4	4	1.00	13	13	0.50	3	6	0.90	9	10
17	42.47	1.00	18	18	1.00	2	2	1.00	3	3	1.00	11	11	1.00	2	2	n/a	0	0	n/a	0	0	n/a	0	0

* session 14 is missing for DZ4B due to a failure of the recording equipment.