K. Stromswold (2007), Genetics and the Structure, Acquisition & Evolution of Language, Biolinguistic Investigations, Santa Domingo, Dominican Republic, February 2007.

Genetics and the Structure, Acquisition & Evolution of Language

Karin Stromswold
Psychology Dept. & Center for Cognitive Science
Rutgers University - New Brunswick

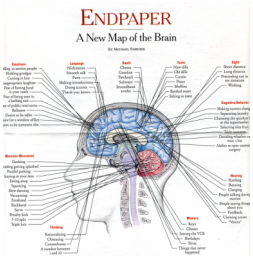
With many thanks to

- **Postdoc**: Ellyn Sheffield (now at Salisbury U.)
- **Lab Manager**: Diane Molnar
- **A small army of research assistants and under graduate students**
- **1400+ parents & children** who participated in the PEGI (Prenatal Environment & Genetic Interaction) project
- NSF (BCS-0446838), Charles & Johanna Busch Biomedical Research Fund, Bamford-Lahey Children's Foundation

Outline of Talk

- Genetics and the neural bases of language
  - Relationship between innateness & heritability
- Genetic factors and language: Some data
  - What we know
    - Linkage studies
    - Twin Studies
  - What we don't know
    - Phenotype-Genotype mapping problem
    - Developmental problem
    - Causal problem: direct vs. indirect effects
    - Interactions problem
    - Specificity problem
- Genetics and the structure of language
- Genetics and the acquisition of language
- Genetics and the evolution of language

Genetics & Biolinguistics



Innateness Hypothesis & Heritability

- **Typical evidence**: Language is universal, learnable, modular
- **Genetic evidence**: If innate predispositions or neural structures enable us to use/acquire language, they **must** be encoded in our DNA
- **Why we might fail to find evidence for language heritability**
  - The Innateness Hypothesis is wrong
  - Environmental factors swamp genetic factors
  - Linguistically-speaking, (normal) people are genetically identical

Heritability = amount of individual variation due to genetic factors

(i.e., if there is no variation, heritability = 0, even if genetic factors determine a trait)

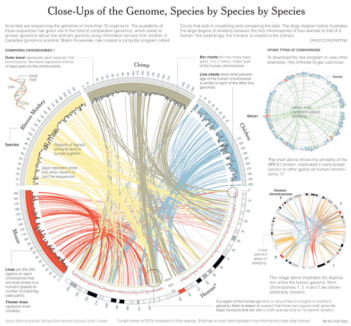
Chomsky (1980): Language is like number of fingers
Lieberman (1984): Language is like height

Innateness Hypothesis & Heritability

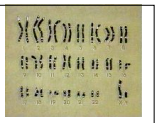
- **Individual differences may exist**
  - **Acquisition rate**: vocabulary (e.g., Goldfield & Reznick, 1990), morphology (e.g., deVilliers & deVilliers, 1973), syntax (e.g., Stromswold 1990, 1995, Snyder & Stromswold 1997)
  - **Adult linguistic proficiency**: verbal fluency (e.g., Day, 1979), compound nouns (e.g., Gleitman & Gleitman, 1970), sentence processing (e.g., Corely & Corley, 1995; Bever et al., 1989), second language acquisition (e.g., Fillmore, 1979, deKeyser, 2000), grammaticality judgments (e.g., Ross, 1979; Nagatu, 1992; Cowart, 1994).
- **Genetic factors may affect impaired but not normal people**
  - Genetic disorders associated with too many/few fingers
  - In contrast with finger length (genetics also play a role in normal finger length)

K. Stromswold (2007), Genetics and the Structure, Acquisition & Evolution of Language, Biolinguistic Investigations, Santa Domingo, Dominican Republic, February 2007.

Genetics of Language: Some Data

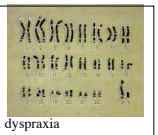


Molecular Genetics of Dyslexia



- **9+ dyslexia loci**
  - 1p34-p36, 2p12-p16, 3p12-q13, 6p21.3-p23, 6q13-q16.2, 7q32, 11p15.5, 15q21-q23, 18p11.2
- **Loci are not specific to types of dyslexia**
  - 15q ≠ orthographic, 6p ≠ phonological
- **Most, if not all, loci are also linked to spoken language disorders**
- **Most, if not all, loci are also linked to other neuropsychological disorders, e.g.**
  - 2p15-p16 schizophrenia
  - 6p21.3-p23: ADHD; schizophrenia
  - 6q13-q16.2: schizophrenia
  - 7q32: autism
  - 11p15.5: ADHD; bipolar disorder
  - 15q21-q23: ADHD
  - 18p11.2: bipolar disorder; schizophrenia
- **Are these bi-linkages merely coincidental** (the loci encompass 1000s of genes)? Or do they indicate that the same genotype may be expressed in different ways?

Molecular Genetics of Spoken Language Impairments



- **FOXP2 gene at 7q31.2**
  - Autosomal dominant disorder with core symptom being speech dyspraxia
  - All and only affected KE family members have aberrant form of FOXP2
  - Not a common cause of spoken language impairment: 10+ studies haven't found it
- **Other spoken language impairment loci exist. E.g.,**
  - **Two markers near FOXP2 gene** (D7S3052, CFT2); **13q21: 16q24; 19q13**:
  - [**Suggestive**: 1p36, 2p15, 6p21, 15q21, 2p22]
  - [**Case reports:** 1p22 and/or 2q31**;** 15q13]
- **Not specific to spoken language impairments. E.g.,**
  - **FOXP2**: Affected people have non-verbal neuropsychological disorders
  - **D7S3052 marker** is near the IMMP2L gene implicated in Tourette syndome
  - **CFT2 marker** is near the WNT2 gene that has been implicated in autism
  - **13q21** has been linked to autism
  - **19q13**: has been linked to autism
  - Most spoken language impairment loci are also linked to written impairments

Behavioral Genetics: Twin Study Rationale

Identical Twins (Monozyogtic, MZ)
- Cotwins share ~ 100% environment
- Cotwins share ~ 100% DNA

Fraternal Twins (Dizygotic, DZ)
- Cotwins share ~ 100% environment
- Cotwins share ~ 50% DNA

⇒ If MZ cotwins are more similar than DZ cotwins, this reflects the role of DNA

⇒ Can quantify the relative importance of DNA and environment by comparing the linguistic similarity of MZ cotwins and DZ cotwins

Genetic Factors & Language Development

- **Twin studies can be used to quantify the role of:**
  - Additive genetic factors (A)
  - Shared environment (C): environment both twins experience
  - Nonshared environment (E): environment only one twin experiences
- **Many twin studies show that genetic factors affect language development**
  - **Meta-analysis (Stromswold, 2001, *Language*)**
    - About 1/2 to 2/3 of language-impaired twins' language
    - About 1/4 to 1/2 of typically-developing (normal) twins' language
  - **PEGI Study (Stromswold 2006, *Cognition*)**
    - Phonology: 68% genetic
    - Lexicon/Lexical Access: 40% genetic
    - Syntax: 59% genetic

Perinatal Environment & Genetic Interactions (PEGI) Study

- N = 267 same sex twin pairs
  - 145 monozygotic (MZ)
  - 122 dizygotic (DZ)
- Age: Between 2 and 6 years old
- Sex: 53% male, 47% female
- Birth Weight
  - MZ: 2294 grams (SD = 583 grams)
  - DZ: 2368 grams (SD = 635 grams)
- Gestational Age at Birth (GA)
  - MZ: 35.1 weeks (SD = 2.8 weeks)
  - DZ: 35.4 weeks (SD = 3.0 weeks)

Linguistic Measures

- **Linguistic Milestones** (parent report)
  - Babble
  - Words
  - Sentences
  - Clear Articulation
- **Language Test Scores**
  - ASQ Communication Scores (all children)
  - For children < 3 years old
    - MacArthur CDI spoken vocabulary (parent report)
    - MacArthur CDI sentence structure (parent report)
  - For children ≥ 3 years old
    - PAL Articulation (imitation of word onsets)
    - PAL Syntax (comprehension of actives, passives, pronouns)
    - PAL Vocabulary/Lexical Access (rapid naming test)
    - PAL Oral Score (PAL Articulation + Syntax + Vocabulary)
- **Speech & Language Therapy/Year**

Non-Linguistic Measures

2 General measures:

Total ASQ score, All Therapies and Special Services/year

7 Gross motor measures:

ASQ gross motor score, 5 milestones (sit, crawl, walk, run, stairs), Physical Therapy/year

5 Fine motor measures:

ASQ fine motor score, 3 milestones (finger feed, fork feed, scribble), Occupational Therapy/year

2 Oral motor measures:

1 milestone (cup drinking), Feeding Therapy/year

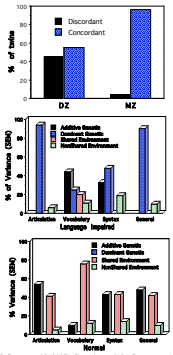
3 Social measures:

ASQ personal-social score, 1 milestone (social smiling), Psych & Behav Therapy/year

2 Cognitive measures:

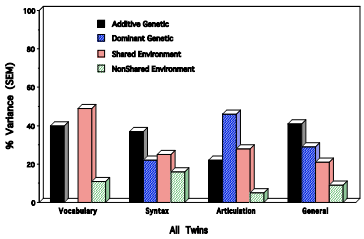
ASQ problem-solving score, Special Education Services/year

Language Impaired & Normal Twins



- Genetic factors account for 82% of language disorders in twins
- Genetic factors account for more of the linguistic variance in language-impaired (LI) than normal twins
- Dominant genetic factors only play a role for LI twins
- For LI and normal twins, genetic factors affect vocabulary the least

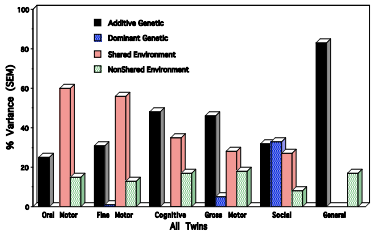
Heritability of Linguistic Abilities: All Twins



Heritable factors account for different amounts of variance for different aspects of language:

Vocabulary < Syntax ≤ Articulation ≤ General

Heritability of Non-Linguistic Abilities: All Twins



Heritable factors account for different amounts of variance for different non-linguistic abilities:

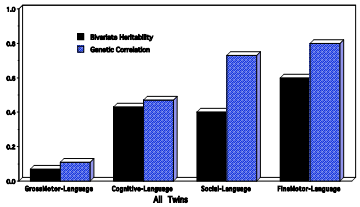
Oral Motor < Fine Motor < Cognitive < Gross Motor < Social < General

Bivariate Analyses

- **Bivariate analyses: Compare twin's performance on test A with cotwin's performance on test B**
  - Bivariate heritability, h_b^2: Extent to which genetic factors account for the phenotypic correlation between test A and test B
  - Genetic correlation, R_g: Extent to which the *same* genetic factors affect A and B regardless of their contribution to the correlation between A and B
- **h_b^2 can be low and R_g high**:
  - e.g., genetic factors play a modest role for A and B, but the same genetic factors affect both A and B
- **h_b^2 can be high and R_g low:**
  - e.g., genetic factors play a substantial role for A & B, but different genetic factors affect A and B

K. Stromswold (2007), Genetics and the Structure, Acquisition & Evolution of Language, Biolinguistic Investigations, Santa Domingo, Dominican Republic, February 2007.

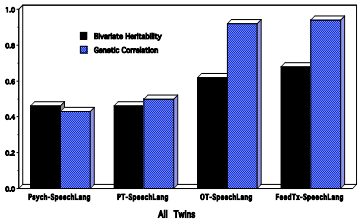
Genetic Overlap Between Linguistic & Non-Linguistic Abilities: ASQ Scores



The amount of genetic overlap between linguistic and nonlinguistic ability differs for nonlinguistic skills

Gross Motor < Cognitive < Social ≤ Fine Motor

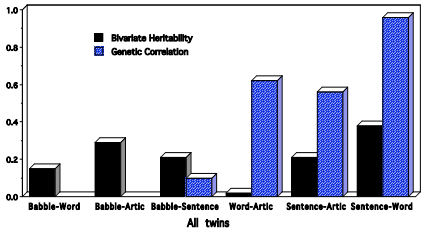
Overlap for Linguistic & Non-Linguistic: Amount of Therapy



The amount of genetic overlap between linguistic and nonlinguistic ability differs for nonlinguistic skills:

(Cognitive) < Psych = Gross Motor < Fine Motor = Oral Motor

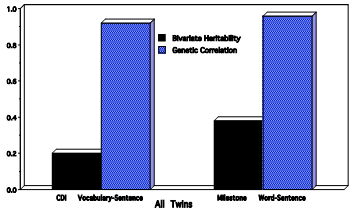
Overlap for Language Milestones



Little genetic overlap for babbling & other language milestones

=> different genetic factors for babbling?

Overlap for Words & Sentences: CDI and Milestone Measures

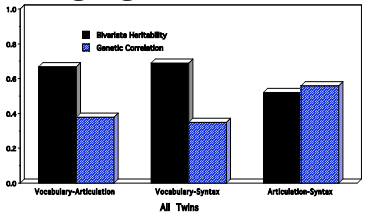


Considerable genetic overlap for words and sentences

⇒**genetic overlap due to both being expressive measures?**

⇒**genetic overlap due to both being parent report measures?**

Overlap for Different Aspects of Language: PAL scores



The genetic overlap between aspects of language differs:

Vocabulary-Syntax = Vocabulary-Articulation < Articulation-Syntax

=> Genetic overlap is *not* just a reflection of expressive vs. receptive measure

=> Genetic overlap a reflection of hierarchical nature of phonology & syntax?

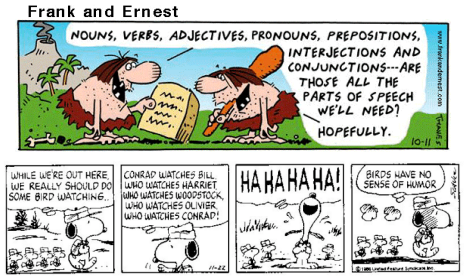
What Isn't Known: Going from Genotype to Phenotype

- **Genotype to phenotype mapping problem**
  - One Genotype:Many Phenotypes (pleiotropy)
  - One Phenotype:Many Genotypes (phenocopy)
- **Developmental problem: phenotypes change**
  - **Compensation**: Initial deficit is compensated for (but compensation may result in a new deficit)
  - **Cascading deficits**: Primary deficit causes additional secondary deficits
- **Causality problem: direct vs. indirect genetic effects**
  - **Direct**: A child has the language disorder "gene(s)"
  - **Indirec**t: A parent with the language disorder gene(s) provides suboptimal input (DLEH)
- **Interaction problem**
  - **Gene-gene interactions**: A child with a language disorder gene or ADHD gene might be fine, but a child with both might be impaired
  - **Environmental interactions**: A child with a mild perinatal brain injury or a mild perinatal hearing loss might be fine, but a child with both might be impaired
  - **Gene-environmental interactions**: A child with a language disorder gene or mild hearing loss might be fine, but a child with both might be impaired
- **Specificity problem**:
  - **Familial Dysautonomia** (9q31; IKBKAP). AR disorder with normal IQs and profound oral motor dyspraxia (but they also have ANS problems)
  - **FOXP2:** Is the language disorder in FOXP2 primary or secondary? What about KE families non-motoric & non-linguistic problems

Common Variant/Multiple Diseases (CV/MD, Becker, 2004)

- **CV/MD:** Common alleles that contribute to a particular disease under certain genetic & environmental conditions can result in a different disease (or no disease) in other conditions
- For a group of related diseases (e.g., autoimmune disorders), some genetic & environmental factors will be unique to a particular disorder and some will be shared by several disorders
- **CV/MD can explain**
  - Why multiple neuropsych. disorders link to the same chromosomal region
  - Why people with the same mutation have different clinical disorders
  - Why simple Mendelian transmission is the exception for familial language disorders
  - Why some language disorder linkage studies fail to find susceptibility loci
  - Why so few genes can cause so many diseases

Genetics & Linguistics



Genetics & Linguistics

- Genetic factors play a non-negligible role in language
- Genetic factors affect normal and language-impaired people' language
- Genetic factors affect all aspects of language
- Some genetic factors are not specific to language
  - Language/oral motor/fine motor overlap
  - Language/social overlap
  - Vocabulary/cognitive
- Some genetic factors are specific to language
- Some genetic factors are specific to parts of language:
  - Genetic factors that influence artic & syntax but not vocab
  - Genetic factors that influence syntax and not articulation

What Genetics can Tell Linguists

- **Language is a biological organ (all parts)**
- **Language is analogous to finger length**
- **Language has non-specific structures**
  - Social/pragmatic structures
  - Complex sensorimotor structures (oral, fine motor, language)
  - Cognitive structures: symbolic representation & meaning
- **Language has language-specific structures**
  - Hierarchical structure?
  - Recursion?

Genetics & Language Acquisition

Language Development & Environmental Factors

- **Twin studies show that genetic factors affect language ….**
- **But they also show that <u>environmental</u> factors play a role**
  - **Meta-analysis (Stromswold, 2001, *Language)***
    - 1/3 to 1/2 of language-impaired twins' language is due to environment
    - 1/2 to 3/4 of normal twins' language is due to environment
  - **PEGI study (Stromswold, 2006 *Cognition*)**
    - Phonology: 32% environment
    - Lexicon: 60% environment
    - Syntax: 31% environment

Genetics & Environment

- **Recall that twin studies can be used to quantify the role of**
  - Additive genetic factors (A)
  - Shared environment (C): environment both twins experience
  - Nonshared environment (E): environment only one twin experiences
- **Problem: twin studies conflate prenatal & postnatal environment**
  - C = shared perinatal + postnatal environment
    - » Shared prenatal = GA, prenatal drugs, intrauterine infection, etc.
    - » Shared postnatal = linguistic input, SES of family
  - E = non-shared perinatal + postnatal environment
    - » Nonshared prenatal = discrepant BW, brain injury, placental blood flow, etc.
    - » Nonshared postnatal = discrepant postnatal accident, illness, etc.

Twins' Language Development

- **Language impairments more common**
  - Twins are 2 to 3 times more likely to be language impaired than singletons
- **Language development lags**
  - Normal twins' language development lags 2 to 3 months behind that of singletons

Why Twins are Delayed: Postnatal Factors?

Generally assumed that twins' linguistic delays are due to suboptimal postnatal environments

- **Less adult linguistic input** (Reznick, 1997; Stafford, 1987; Tomasello, Mannle, & Kruger, 1986; Conway et al., 1980; Lytton, Conway, & Suave, 1977; Thorpe, Rutter, & Greenwood, 2003)
- **Less complex adult linguistic input** (Conway et al., 1980; Thorpe, Rutter, & Greenwood, 2003)
- **Suboptimal non-linguistic environment:** predict linguistic and cognitive delays, but not motor delays

Why Twins are Delayed: Perinatal Factors?

- **Prematurity:** < 37 weeks gestation age at birth (GA)
  - Twins are 5 times more likely to be born prematurely
- **Low Birth Weight:** < 2500 grams at birth (~5.5 lbs)
  - Twins are 10 times more likely to be low birth weight
- **Perinatal Complications**
  - Prenatal complications: amniotic, placental, infection, etc.
  - Intrapartum complications: cord, breech, hemorrhage, etc.
  - Postnatal complications: severe jaundice, BPD, NEC, sepsis
  - > Brain injuries: IVH, PVL, hypoxic/ischemic injuries etc.
- **Perinatal hardships are associated with**
  - Smaller brains
  - Global delays/impairments (linguistic, cognitive, and motor)

Perinatal vs. Postnatal Environment

Conway et al. (1980): Maternal speech affects twins' language more than perinatal factors

- **Twins' and singletons' data combined in analyses**
  - » Thus, study says nothing about the twin-singleton language gap
- **Study is flawed**
  - » Very small study (only 12 twin pairs)
  - » Low risk twins (mean GA = 38 wks; mean BW = 2800 g)
  - » Little variance for perinatal measures
  - » Large variance for postnatal measures

Perinatal vs. Postnatal Environment

Thorpe et al. (2003): Maternal speech accounts for more of the twin-singleton language gap than perinatal factors. Problems:

- **Eliminated high-risk children (< 33 weeks, brain injured or disabled)**
- **Maternal speech composite measure is optimized to account for the gap:**
  - Eliminated postnatal variables that affected twins' and singletons' language equally
  - Examined many postnatal variables and chose the best 6 predictors of language

Rutter et al. (2003): Perinatal factors account for none of the gap.

- **Eliminated high-risk children (< 33 weeks, brain injured or disabled)**
- **Perinatal composite measure is a poor measure of perinatal risk**
  - Included factors that don't affect outcome (1 min. Apgar; perineal lacerations)
  - Excluded factors that do affect outcome (hypoxia, infection, jaundice)
  - BW & GA were normalized separately for twins & singletons (which assumes conclusion)

Teasing Apart Perinatal vs. Postnatal

- **MZ and DZ cotwins share ~100% of their environment**
  - Thus, if environment affects a trait, it will increase phenotypic similarity of MZ and DZ twin pairs equally
- **The result:**
  - Role of shared environmental factors (C) will increase
  - Role of genetic factors (A) will decrease
- **If delays are due to adverse perinatal environment**
  - High risk perinatal: Higher C & lower A than low-risk perinatal
- **If delays are due to adverse postnatal environment**
  - High risk postnatal: Higher C & lower A than low-risk postnatal

Teasing Apart Perinatal vs. Postnatal

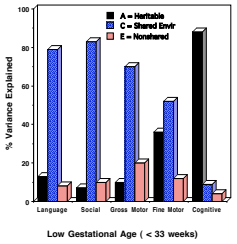
- **Adverse perinatal environment (low GA) overshadows genetic factors in verbal & nonverbal development** (Koeppen-Schomerus et al., 2000)
  - GA < 32 weeks: 100% variance due to shared perinatal environ
  - GA ≥ 32 weeks: 18-32% variance due to genetic factors
- **Adverse postnatal environment (low SES) overshadows genetic factors in full scale IQ** (Turkheimer et al., 2003)
  - Low SES: 10% genetic, 58% shared postnatal environment
  - High SES: 72% genetic, 15% shared postnatal environment
- **Can't tell whether peri- or post-natal environment is more important because different twins studied**

The Role of Prenatal & Postnatal Environment Can Differ for Different Things

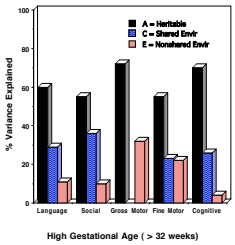
Prenatal Environment
Twin-Twin Transfusion
(donor twin on right)

Postnatal Environment
Parenting
(parent on left)

Prenatal: Low vs. High GA



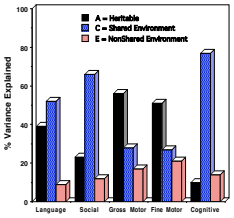
Low Gestational Age (< 33 weeks)



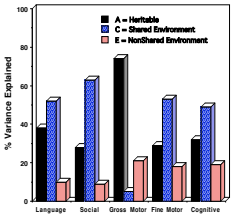
High Gestational Age (> 32 weeks)

- Perinatal hardship depresses heritability for language but *not* cognition.
- Perinatal hardship inflates shared environment for language but *not* cognition.

Postnatal: Low vs. High SES



Low Socio-Economic Status



High Socio-Economic Status

- Postnatal hardship depresses heritability for cognition but *not* language.
- Postnatal hardship inflates shared environment for cognition but *not* language.

Language Acquisition & the Role of the Environment

- **Studies with perinatally high risk twins**
  - Underestimate the role of genetic factors (A)
  - Overestimate the role of shared environmental factors (C)
- **Perinatal shared environment affects language**
  - Perinatal factors are mainly <u>biological</u> (which affect neurodevelopment)

    ==> Therefore, neurobiological factors affect language
- **Postnatal shared environment affect cognition**
  - Postnatal factors are mainly <u>psychosocial</u> (which include SES, adult input, etc.)

    ==>Therefore, psychosocial factors affect cognition

K. Stromswold (2007), Genetics and the Structure, Acquisition & Evolution of Language, Biolinguistic Investigations, Santa Domingo, Dominican Republic, February 2007.

What Genetics can Tell Language Acquisitionists

Language acquisition is a biological program unfolding (akin to puberty)

Genetic findings support acquisitional theories that posit

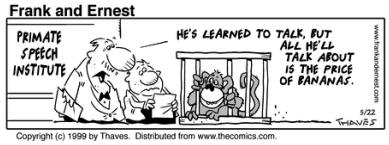
1. **Innate, general biological mechanisms**: Non-specific genes
2. **Innate, language-specific mechanisms**: Language-specific genes

Environmental findings support

1. **Nativist theories of language acquisition:**
   Children's neural status > Adult linguistic input
2. **Empiricist theories of cognitive development:**
   Adult input > children's neural status

Genetics & Language Evolution

Copyright (c) 1998 by Thaves. Distributed from www.thecomics.com.

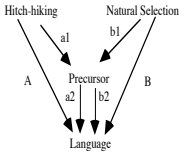


Copyright (c) 1999 by Thaves. Distributed from www.thecomics.com.

Genetics & Language Evolution

- ***A*** = amount of individual variation due to genetic factors
- Significant ***A*** for language means that genetic variance exists *now*
- Genetic variance for language means something that contributes to linguistic ability could have been selected for.
  - It does not mean that it *was* selected for
  - Even if it was selected for, it does not say *what* was selected for
- Genetic overlap for linguistic & nonlinguistic abilities is consistent with them having shared evolutionary history
- Genetic overlap for subcomponents of language is consistent with them having shared evolutionary history

Theories of Language Evolution



- Process by which genetic factors that subserve language came into existence (serendipity vs. natural selection)
- Were any of the genetic factors specifically linguistic or did they all piggy-back on other abilities (communication, oral motor, fine motor (object manipulation), cognition, mental representation?
- If some were specifically linguistic, what were their functions?

Genetics & Language Evolution (1)

1. **Genetic factors play a substantial role for LI & normal twins' language**
   ⇒ Some language-related ability may have been selected for
2. **Genetic factors play a greater role for LI than normal twins' language**
   ⇒ Different genes?
   ⇒ Different evolutionary history?
   (NB: Currently no "stoppage" with language-impairment)

Genetics & Language Evolution (2)

3. **Genetic factors affect all aspects of language, but articulation & syntax are affected more than vocabulary**
   ⇒ Greater selective pressure for articulation & syntax?
   ⇒ More variance for articulation & syntax?
4. **Genetic factors also play a substantial role for all nonlinguistic abilities (with the possible exception of oral motor in normal twins).**
   ⇒ Role greater for LI twins because their language impairment is non-specific?

K. Stromswold (2007), Genetics and the Structure, Acquisition & Evolution of Language, Biolinguistic Investigations, Santa Domingo, Dominican Republic, February 2007.

Genetics and Language Evolution (3)

5. **Genetic overlap for linguistic & non-linguistic skills**

   Language and Fine Motor: Shared evolutionary history?
   Language and Oral Motor: Shared evolutionary history?
   Language and Social Skills: Shared evolutionary history?

6. **Genetic overlap is greatest for syntax and articulation**

   ⇒ Shared evolutionary history due to their hierarchical structure?

7. **Small amount of syntax-specific genes**

   ⇒ Recursion?

Implications of Genetic Results

1. **Genetic findings support biological basis of language**
   - Some neural circuitry is quite general (e.g., sensorimotor)
   - Some neural circuitry is quite specific (e.g., articulation/syntax)
2. **Genetic findings support linguistic theories that posit structures that are**
   - General structures: social/pragmatics, sensorimotor, cognitive
   - Language-specific structures: Hierarchical structure? Recursion?
3. **Genetic findings support nativist theories of language acquisition that posit**
   - Innate, general mechanisms: Non-language specific genes
   - Innate, language-specific mechanisms: Language-specific genes
4. **Genetic findings support evolutionary theories that posit different evolutionary histories for**
   - Non-language specific abilities (FLB)
   - Language-specific abilities (FLNarrow-ish: recursion + hierarchical structures)

Thank you!