جامعة نيويورك أبوظبي

NYU | ABU DHABI

# PSYCH-UH 2218: Language Science

## Class 18: The logical problem of language acquisition

Prof. Jon Sprouse
Psychology

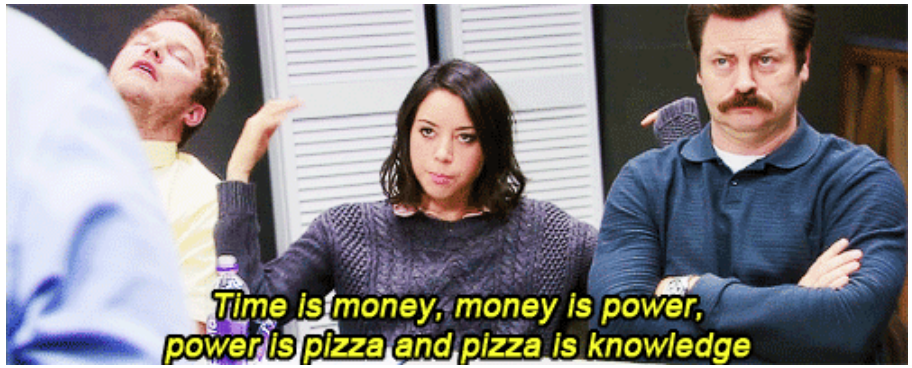
Fact 1: To learn a language means to learn the rules that generate a very large, probably infinite, set

# The infinity of language means that we must learn the intention of the set, not extension

| **intension** | **extension** |
| --- | --- |
| XP → (WP) X’ | Sarah wrote a novel. |
| X’ → (ZP) X’ (ZP) | Lisa claims that Sarah wrote a novel. |
| X’ → X (YP) | Mary thinks that Lisa claims that Sarah wrote a novel. |
| | … |

It is obvious that the extension of an infinite set cannot be memorized by a finite mind. There are ~80 billion neurons in the brain. That is a lot. But even if we only needed one neuron to memorize a sentence (unlikely!), it is still far smaller than infinity.

Beyond that, we can understand sentences we have not heard. So we know that they are not memorized (because memorization requires hearing it at least once)!



# To be clear, it is not just syntax that is very large and probably infinite

syllable → onset rhyme

rhyme → nucleus coda

onset → (C)(C)(C)

nucleus → V

coda → (C)(C)(C)(C)

phonotactic constraints

| **existing words** | **possible words** |
|---|---|
| trim | twim |
| twin | trin |
| | … |

Phonology and Morphology are also characterized as sets of rules that generate sequences that we might call words. Yes, the number of words that any given language uses is finite. But the set of **possible words** is very large, and probably infinite.

You can see this because new words are created all the time. And speakers can provide judgments about which sequences could be new words, like **trin**, and which sequences could not, like ***tlin**.

# Fact 1: To learn a language means to learn the rules that generate an infinite set

XP → (WP) X’

X’ → (ZP) X’ (ZP)

X’ → X (YP)

movement

theta criterion

…

syllable → onset rhyme

rhyme → nucleus coda

onset → (C)(C)(C)

nucleus → V

coda → (C)(C)(C)(C)

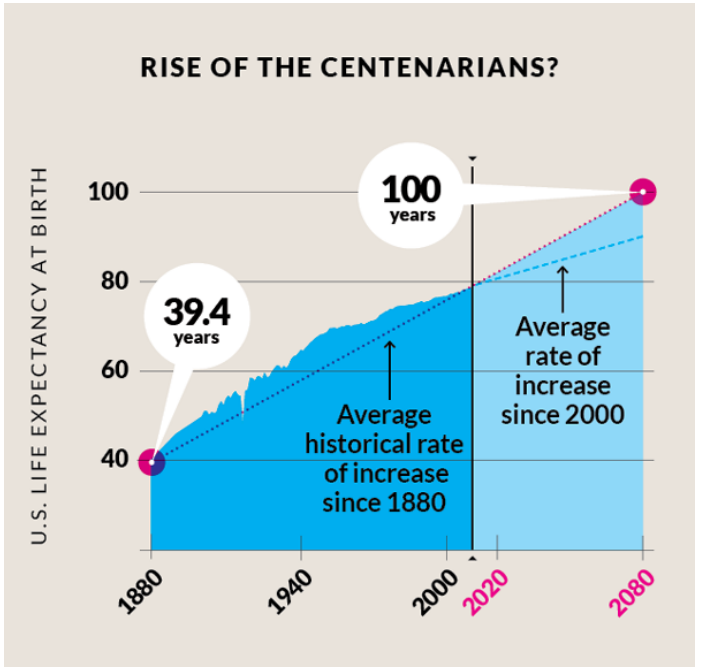
phonotactic constraints

I know I keep harping on this, but this is a critical fact for understanding the logical challenge posed by language. This is the **end state** of language acquisition. It is **the goal**. If someone has successfully learned a language, what they have learned is not a list of sentences, it is the **rules that generate that set**.

Fact 2: The evidence that children receive about those rules is **finite**

# The human lifespan is finite

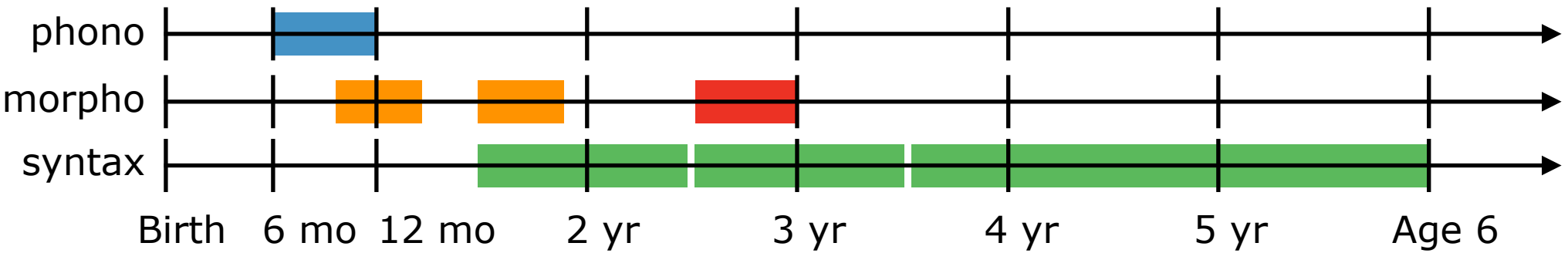
I am sorry to always remind us of our own mortality in this class, but it is a critical fact. Because our lifespan is finite, **we only hear a finite number of sentences in our lives**!




This is a good thing though. It means we can still laugh even when we get old. If it were possible to hear everything in our lives, we'd run out of jokes!

# The timeline of language acquisition is actually fairly condensed

For children that are exposed to a language from birth (in the absence of child abuse, or moving, etc), the process of learning phonology, morphology, and syntax tends to be complete by about age 6:




We will look at this in more detail next week. For now, I just want to observe that children complete language acquisition in about 6 years. So they only hear as many sentences as can be heard in 6 years. (Hart & Risley 1995 estimate that children hear around 2 million utterances in those 6 years.)

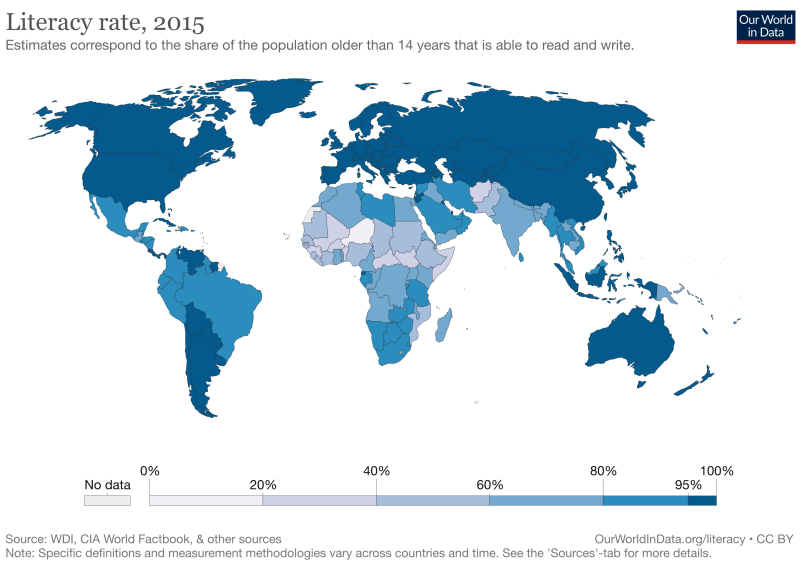
This is another critical point. The evidence that children receive (hearing utterances) for language acquisition is **finite**.

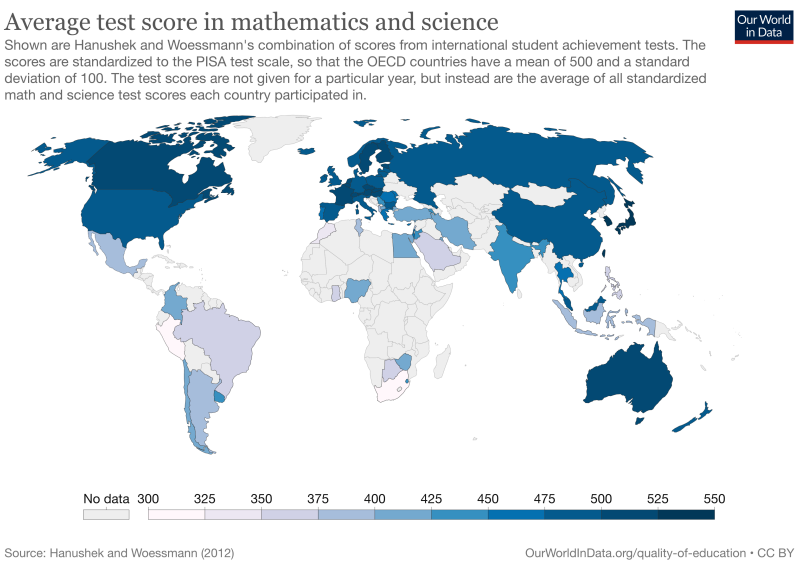
Fact 3: In the absence of disease/ disorder/abuse, everyone is equally successful at learning their first language

# What do we mean by equally successful?

First, everybody learns language. In the absence of disease or child abuse, we don't find humans who simply fail to learn language.

Compare this with most other complex skills:




Or other activities, like playing a musical instrument, ice skating, dancing, etc.

The typical expectation is that some humans will learn a skill, and others won't. But for language **everybody learns it.**

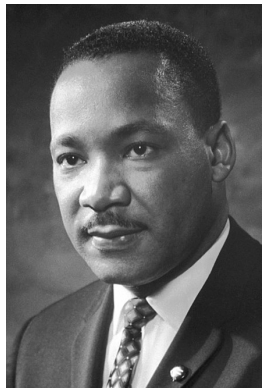
# What do we mean by equally successful?

Second, everybody seems to learn language to the same degree.

For two speakers who are both native speakers of the same language, there is no way in which one is better than the other at speaking the language.

But, wait, people do say that some people are "better" at language than others?

Maya Angelou



Martin Luther King Jr.

But when we say this, we mean something like an artistic skill that uses language, not language itself. Like Maya Angelou's ability to write poetry and prose, or MLK's ability to write and deliver a speech. (Sometimes we also mean an adult's ability to learn multiple languages later in life — but we will discuss that next week.)

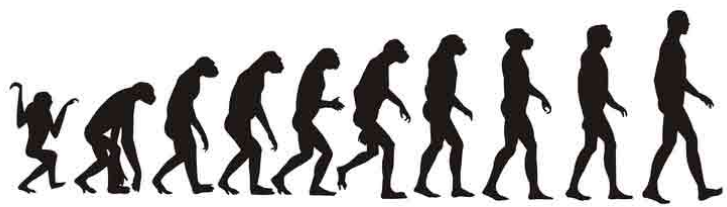
At the level of grammar, everyone has the same ability. We all learn the same set of rules, and we can all use them equally. Nobody is like "Jon can form sentences that I can't form" Or "Jon can use allophones that I can't use". Unlike other complex skills, **we all attain the same level of proficiency**.

# An interesting comparison

Language is complex, so it feels like it is similar to complex tasks like math, music, reading, etc.

But, the fact that everyone attains the same level of proficiency in language makes it more similar to a number of cognitive abilities that we would probably characterize as **simpler.** But this simplicity is **an illusion**. The real issue is that these are **fundamental abilities of our species:**

**Walking upright**. Every human does this (in the absence of disease or abuse). But it is a complex ability.



**Vision**. Every human does this (in the absence of disease or abuse). But the development of vision is a complex process.


**Memory**. Every human has a complex memory system (in the absence of disease). But, it also develops through complex processes.

# Fact 4: The challenge of learning the rules of an infinite set from a finite subset

# Let’s try it with numbers

I have a single rule in mind that can be used to generate a sequence of 3 numbers. This one rule can generate an infinite number of sequences (because numbers are infinite!)

Here is a sequence that is generated by the rule: **2, 4, 8**

What is the rule?

Take a moment to make a hypothesis. What do you think the rule is?

And now think about how you could test your hypothesis. What information would help you figure out if your rule is correct.

This is called the Wason 2-4-6 task (Wason 1960). Peter Wason was a psychologist who studied how humans use logic to solve problems. You may have encountered this in another class (or other Wason tasks, like the Wason selection task). His point is that humans are not very logical. But our point is that **learning the rule that generates an infinite set from a finite set is very difficult.**

# Let’s try it with numbers

Here is a video of some people being confronted with the task. Let’s watch it to see what works and what doesn’t.

https://www.youtube.com/watch?v=vKA4w2O61Xo&t=262s&ab_channel=Veritasium

In the Wason 2-4-6 task, participants are told that they can state new sequences that fit their rule, and the experimenter will tell them if it matches the rule or not. That is what will happen in this video.

# Two types of evidence

**Positive Evidence:** Evidence about which items are **present** in the infinite set.

**Negative Evidence:** Evidence about which items are **absent** from the infinite set.

When the interviewer says yes, that is positive evidence. And when the interviewer says no, that is negative evidence.

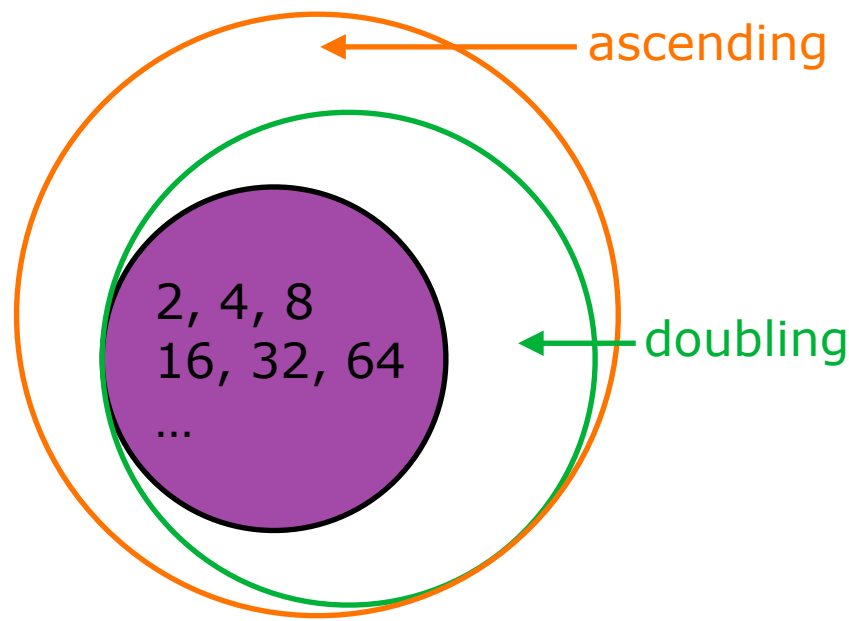
# It is critical to receive **both** types of evidence

**Positive Evidence:** Evidence about which items are **present** in the infinite set.

**Negative Evidence:** Evidence about which items are **absent** from the infinite set.

The problem with positive evidence **alone** is that a finite number of observations (finite evidence) is compatible with an infinite number of theories. (In philosophy this is called the problem of induction.)

With only positive evidence, you can **get stuck** on an incorrect hypothesis that happens to be a subset of the correct hypothesis.




Positive evidence in language acquisition is hearing other people's utterances. Children **definitely have access to this.** Here is a project called CHILDES that builds corpora of sentences spoken to and around children: http://childes.psy.cmu.edu/

# It is critical to receive **both** types of evidence

**Positive Evidence:** Evidence about which items are **present** in the infinite set.

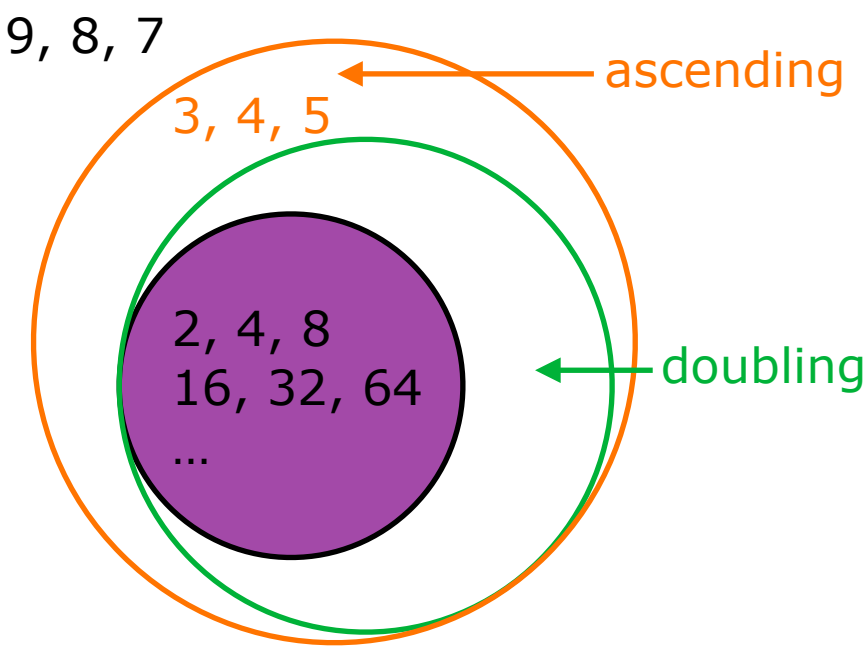
**Negative Evidence:** Evidence about which items are **absent** from the infinite set.

If we add the possibility of negative evidence into the system, you can then **test hypotheses**:

You can produce a sequence that matches ascending but not doubling, like 3, 4, 5.

And you can test ascending by producing a sequence that doesn't match ascending, like 9, 8, 7.

But notice that this **requires that someone could tell you NO, i.e., give you negative evidence.**



# Fact 5: Children do not receive unambiguous negative evidence

# There is no explicit instruction during language acquisition

The obvious source of negative evidence for children would be **explicit instruction**. But children do not receive explicit instruction during language acquisition. I think we all know this intuitively. But let's be systematic.

First, language acquisition is mostly complete by age 6, which is when most formal education begins. Children can already talk when they go to school. This is a prerequisite for learning everything else in school! So school cannot be the source of language.

Second, even when language is discussed in school (or by parents at home), we don't really cover everything that the child would need to learn the language.

fanfreakingtastic

absofreakinglutely

Finally, linguistics as a field exists because speakers of a language **don't consciously know the rules of their languages**. If speakers (and teachers, parents, etc) already knew all the rules, we wouldn't need to do this work. And we certainly wouldn't **still be trying to figure out mysteries**. The fact that professional scientists in the 21st century still can't figure everything out means that there is no way that children are being taught language explicitly.

# What could negative evidence look like?

There are ways other than explicit instruction for children to receive negative evidence. Let's say a child has a hypothesis in mind about the rules of the grammar. The child will then use those rules to generate their own utterances.

Negative evidence in language could be some sort of response by the parent after a child produces a sentence that does not follow the parent's rules of grammar. In other words, some sort of signal after the child produces an ungrammatical sentence that tells the child it was ungrammatical.

This response need not be an explicit correction. It could take any number of forms (these are adapted from Marcus 1993):

| | |
|---|---|
| **Explicit disapproval:** | Parent says no or shakes head. |
| **Non sequiturs:** | Parent fails to understand the child. |
| **Repetitions:** | Parent repeats the child's utterance. |
| **Recasts:** | Parent corrects the child's utterance. |
| **Questions:** | Parent asks for more information. |

# But children ignore this feedback

**child:** Want other one spoon, Daddy.

**parent:** You mean, you want the other spoon.

**child:** Yes, I want other one spoon, please Daddy.

**parent:** Can you say “the other spoon”?

**child:** Other… one… spoon.

**parent:** Say “other”.

**child:** Other.

**parent:** “Spoon”.

**child:** Spoon

**parent:** “Other spoon”.

**child:** Other… spoon. Now give me other one spoon?

# Or children misinterpret it

**child:** Nobody don’t like me.

**parent:** No, say “nobody likes me”.

**child:** Nobody don’t like me.

**parent:** No, say “nobody likes me”.

**child:** Nobody don’t like me.

**parent:** No, say “nobody likes me”.

**child:** Nobody don’t like me.

**parent:** No, say “nobody likes me”.

**child:** Nobody don’t like me.

**parent:** No, say “nobody likes me”.

**child:** Oh! Nobody don’t likes me.

# Feedback is actually **ambiguous**

A study by Bohannon and Stanowicz 1988 looked at the feedback that parents provide to children. They found that parents provide feedback (of the types listed earlier) after **both** ungrammatical sentences and grammatical sentences. They found that parents gave feedback to children after ungrammatical sentences 35% of the time; and they gave feedback to children after grammatical sentences 14% of the time.

Think about that. If parents are giving feedback after both types of sentences, children can't use that feedback to identify ungrammatical sentences. In other words, feedback is **noisy**. It is not a clear indicator of ungrammaticality.

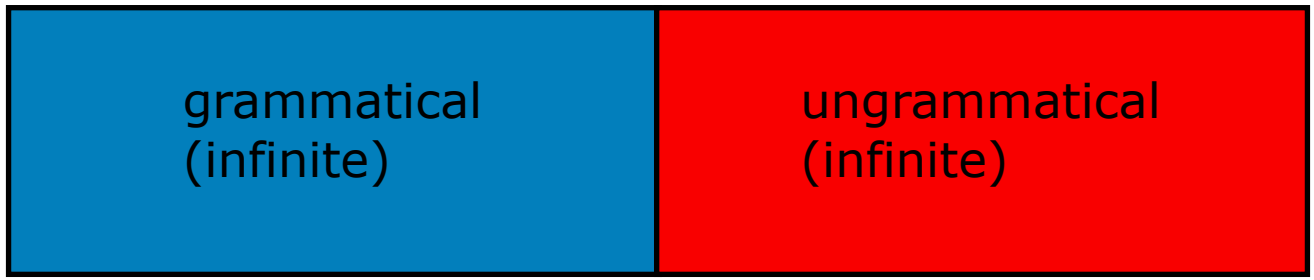
I won't go into the math, but Marcus 1993 calculated that the rates of feedback for ungrammatical and grammatical sentences mean that **children would have to repeat a sentence 85 times** in order to determine whether the feedback that they were receiving was because it was ungrammatical, or whether it was because it was grammatical (i.e., to figure out if it is the 35% rate or 14% rate). Obviously, children don't repeat sentences 85 times to figure out if they are part of the language or not.

# Maybe non-occurrence is negative evidence?

It is tempting to say that children do receive negative evidence in the form of a sentence never occurring. They do hear 2M sentences by age 6. So, if they fail to hear a sentence, perhaps they can use that as negative evidence.

But we have to remember that there are an infinite number of sentences that children never hear. One subset of those sentence is part of the language — that subset is infinite; and the other subset is not part of the language — it is also infinite. How do children figure out which is which?

**The set of unheard sentences (infinite)**




Time is money, money is power, power is pizza, pizza is knowledge.

*What do you wonder whether Lisa invented?

Remember, the challenge is that children have to figure out which strings they haven't heard because of chance, and which they haven't heard because the sentence is ungrammatical.

# The logical problem of language acquisition

# The logical problem of language acquisition

And now we are ready to lay out the logical problem of language acquisition:

**Fact 1:** Learning human language is learning the rules that generate a very large, probably infinite, set.

**Fact 2:** The evidence that children receive about those rules is finite.

**Fact 3:** In the absence of disease or abuse, all children succeed in learning language, and all succeed to the same degree.

**Fact 4:** Learning the rules that generate an infinite set from finite evidence alone requires both positive and negative evidence.

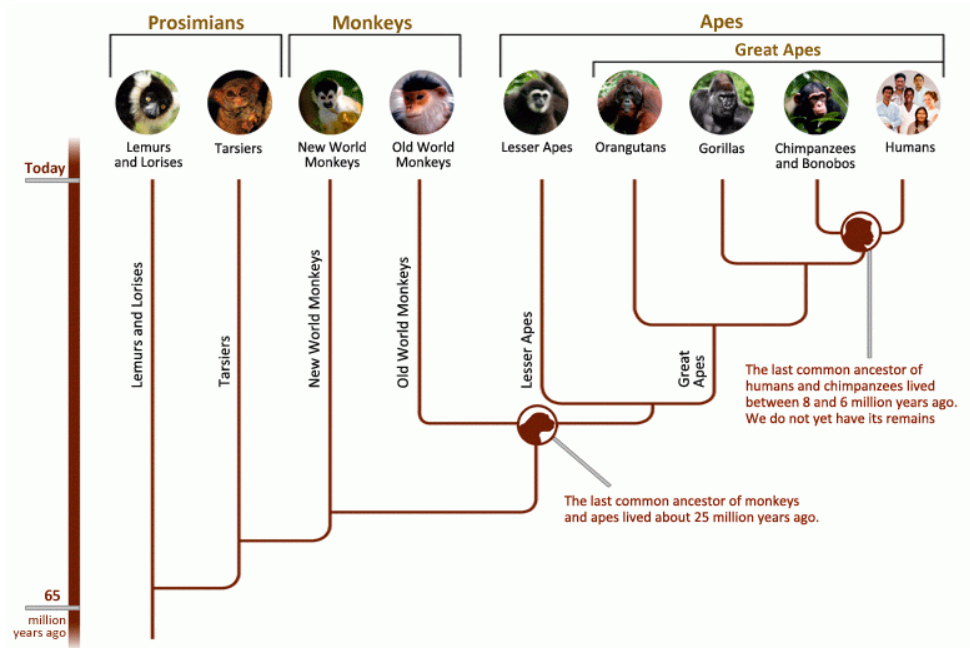
**Fact 5:** Children do not receive (or make use of) negative evidence.

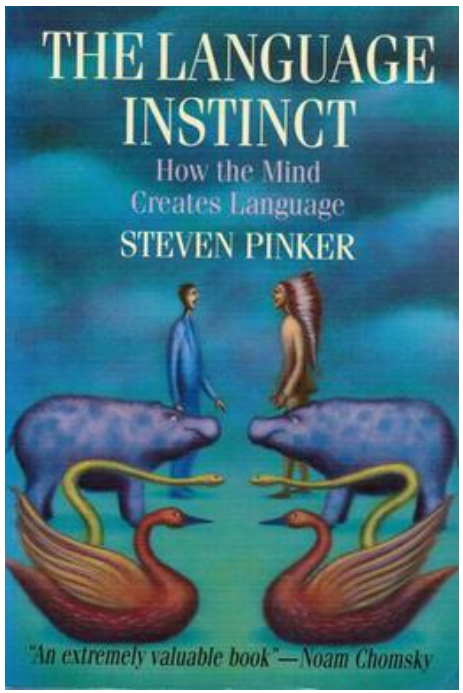
**Conclusion:** Children must have some other mechanism that ensures that all children successfully learn language (the same way all learn to walk, see, etc).

# The genetic hypothesis (also called Nativism)

# Language is obviously part of our genetic endowment as humans

We are the only animal species with language. Other animals have communication systems (and we will study them in detail in a couple of weeks!). But none have phonology, morphology, and syntax the way that we do. Even primates can't do what we can.



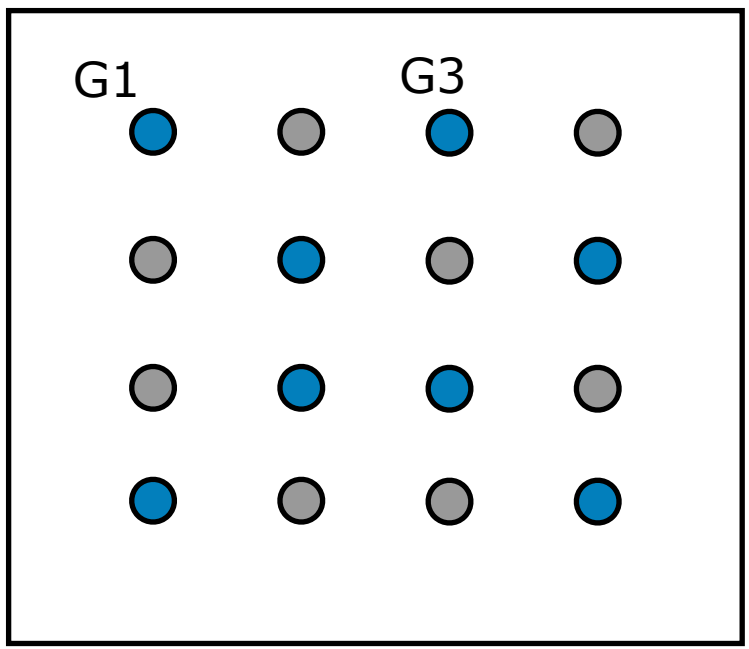




Much like walking upright or vision, the ability to have language seems more like an "instinct" - something that is part of the genetic endowment of all humans.

# Part of that endowment must be mechanisms that ensure successful acquisition

Again, this is obviously true. What is the point of a genetic endowment for an ability if it does not guarantee successful development of that ability?

But the logical problem of language acquisition makes it clear to us what those mechanisms must accomplish:




Hypothesis space

All learning requires a **hypothesis space** - the set of all possible hypotheses that the learner could entertain.

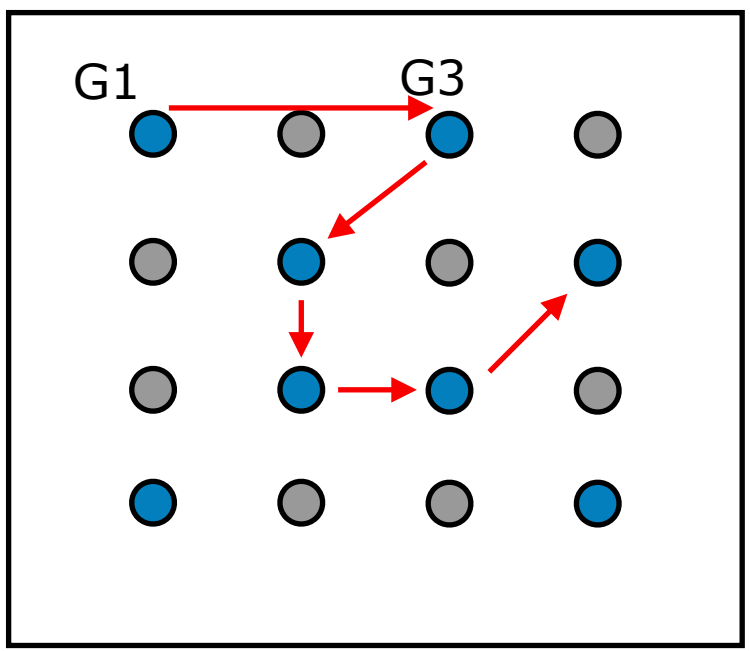
We can view this as a space of possible grammars that the child could hypothesize.

Part of the genetic endowment of language will be the fact that some grammars are possible and some are not. For example, a syntactic rule that doesn't follow X-bar theory is probably not possible. A phonology built on amplitude is probably not possible. This simplifies the task by reducing the number of hypotheses

# Part of that endowment must be mechanisms that ensure successful acquisition

Again, this is obviously true. What is the point of a genetic endowment for an ability if it does not guarantee successful development of that ability?

But the logical problem of language acquisition makes it clear to us what those mechanisms must accomplish:




Hypothesis space

All learning requires an algorithm for evaluating one hypothesis and adopting a new hypothesis based on evidence.

We can think of this as moving through the hypothesis space based on the evidence that children receive (positive evidence only).

Part of the genetic endowment for language will be a mechanism that prevents children from "getting stuck" when they use positive evidence, and ensures that all children succeed in language acquisition.

# How do we study this?

This is a big endeavor. We have only really begun to scratch the surface of this, even after 70+ years of doing it with modern scientific tools.

One step is to compare the grammars of all human languages. Anything that varies between languages must be learned from evidence. And anything that is common across languages could possibly be part of the genetic endowment.

Another step is to compare communication systems across species. Anything that is common across species is part of a broad genetic endowment, and anything that is specific to humans could possibly be part of our genome.

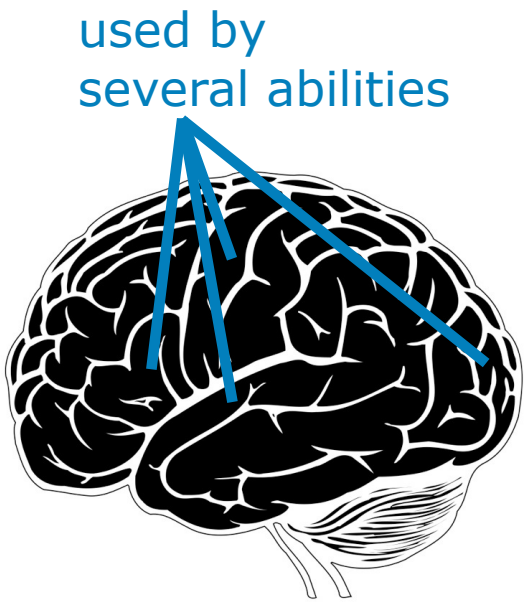
Another step is to explore what can be learned from positive evidence in principle — anything that cannot be learned from positive evidence is possibly part of the genetic component.

Another step is to explore the grammars that children hypothesize at different ages. This will help us to see how they move through the hypothesis space. We can see which hypotheses they adopt, which hypotheses they do not adopt, and perhaps even how they move from one to the next.

# An extra dimension

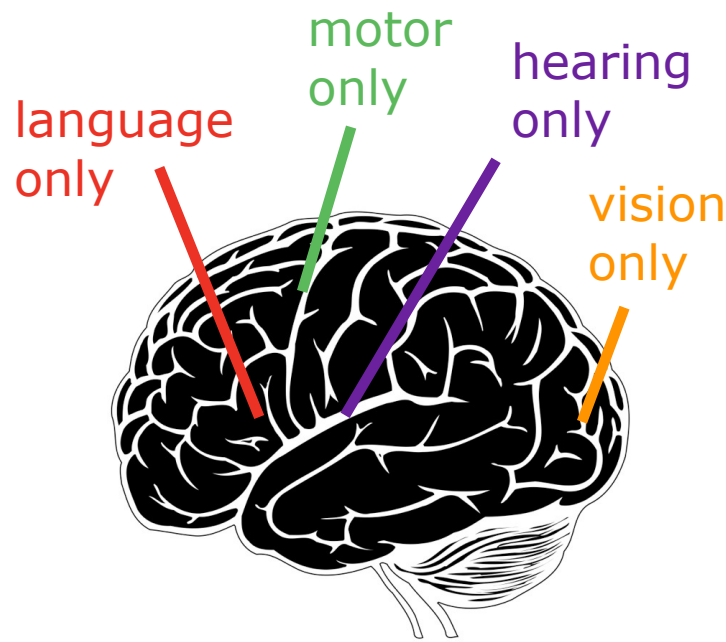
There are two possible types of mechanisms that could be part of our genetic endowment:

Domain-general mechanisms are used by multiple cognitive abilities.




Tracking probabilities may be domain-general.

Domain-specific mechanisms are used by one cognitive ability.




Something like X-bar theory might be domain-specific.

As we uncover potential mechanisms that are part of the genetic endowment, we can also ask whether they are domain-general or domain specific.