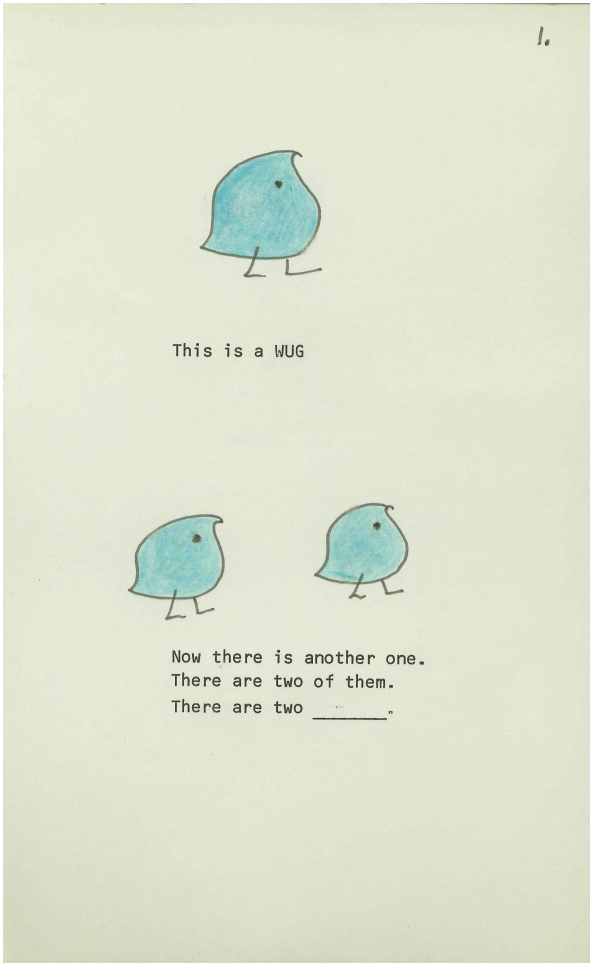
It’s the exceptions that prove the rule or when & why Adam first “feeled dat”

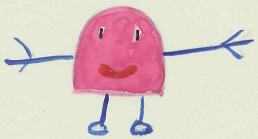
6.S077 Recitation 8 (Week 9)

The drive to learn rules

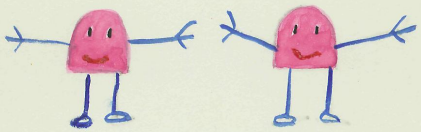


Berko, 1957: “Wug” test

How much do children “over-regularize”?



This is a KAZH.

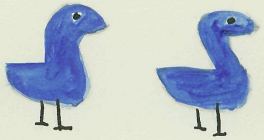


Now there is another one.

There are two of them.

There are two _____.

This is a GUTCH.

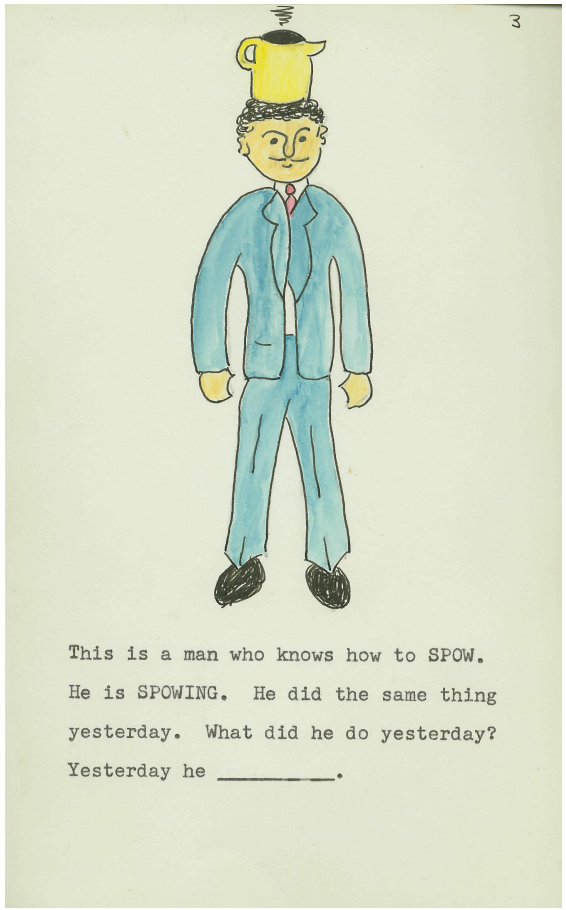


Now there is another one.

There are two of them.

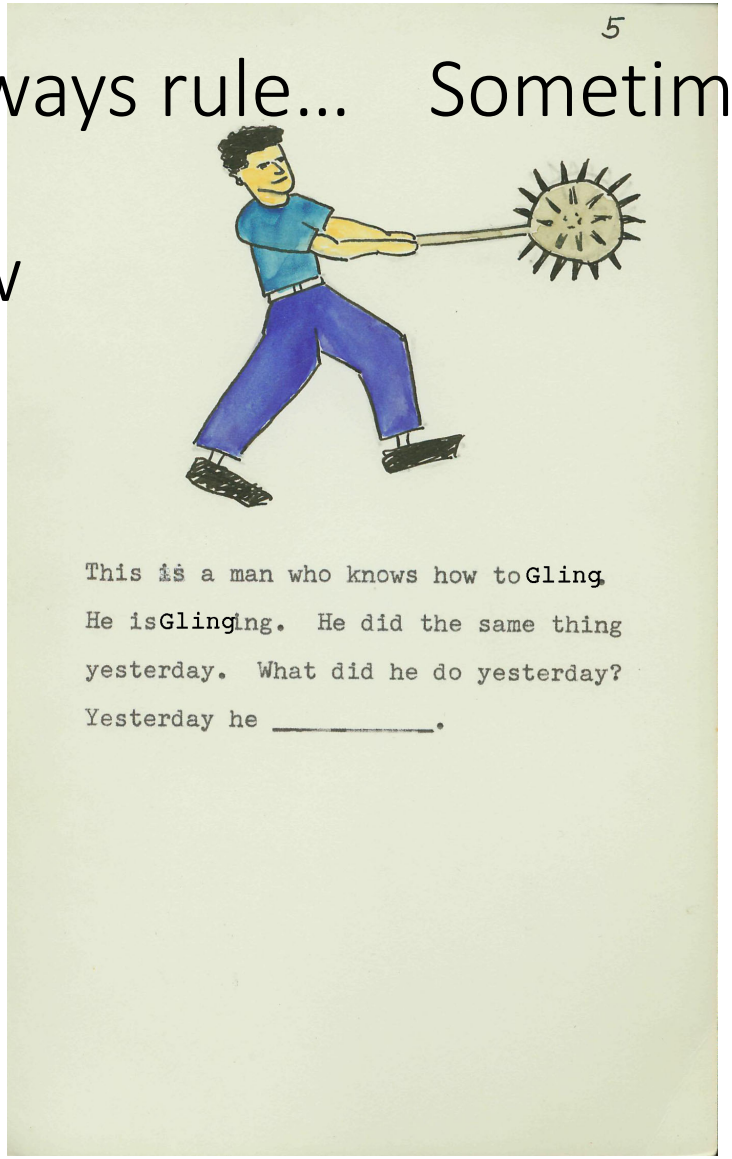
There are two _____.

Rules are inevitable due to the infinity of language and the sparsity of examples



But rules don't always rule… Sometimes there are exceptions
Why? When? How to generalize?

Only 1 child in 86 said "Glung"
The other children were quiet….



gentif mawg

gentif lepal

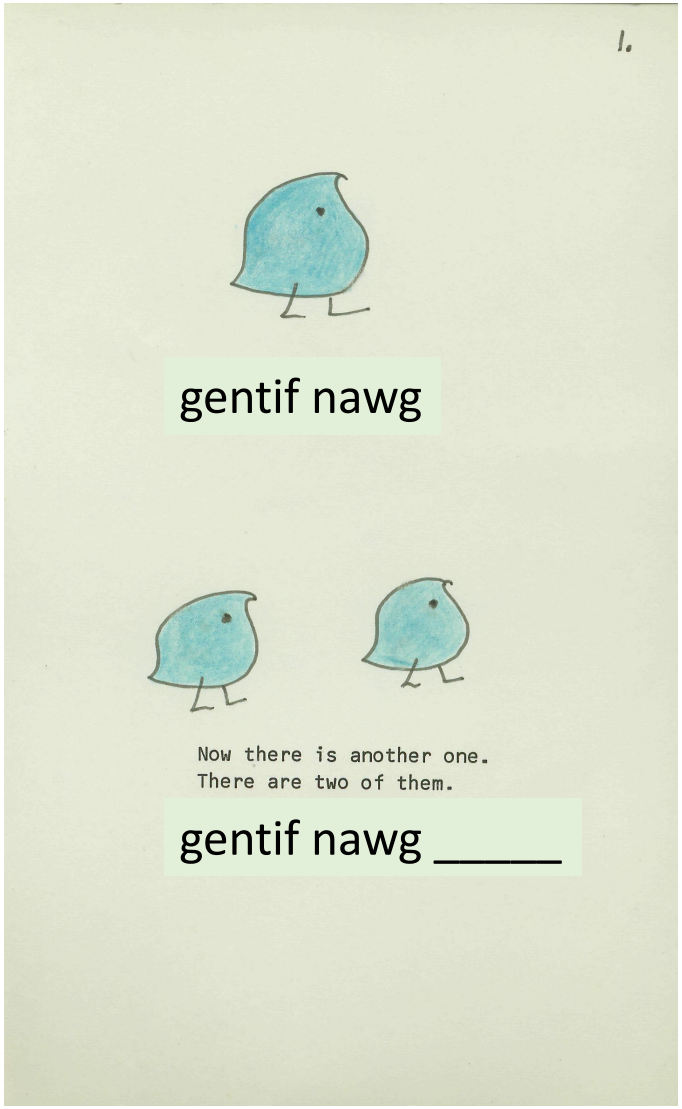
gentif mawg ka

gentif tomber ka

gentif tomber

gentif lepal

gentif lepal po



We have already seen:
2 regular examples
1 “exception”

Will a child generalize or not?
What about you?

When do kids form rules? Allow exceptions?

- Must allow exceptions! "All grammars leak"
- Example:

  Do you have any wool?

  Don't ask what you can ...
- English since about 1680: use do (and don't put the verb in front)
- But, there are some common examples (even in child input) that still put the verb at the front):

  Baa, baa, black sheep have you any wool?

  Ask not what you can do for your country …
- But how many exceptions are allowed before we give up and memorize everything as a list?

Case study: regular/vs irregular verbs & past tense

- English regular past tense for verbs: add “d”:

  bake-baked; cook-cooked; feel-felt; …; walk-walked; zoom-zoomed (google-googled)

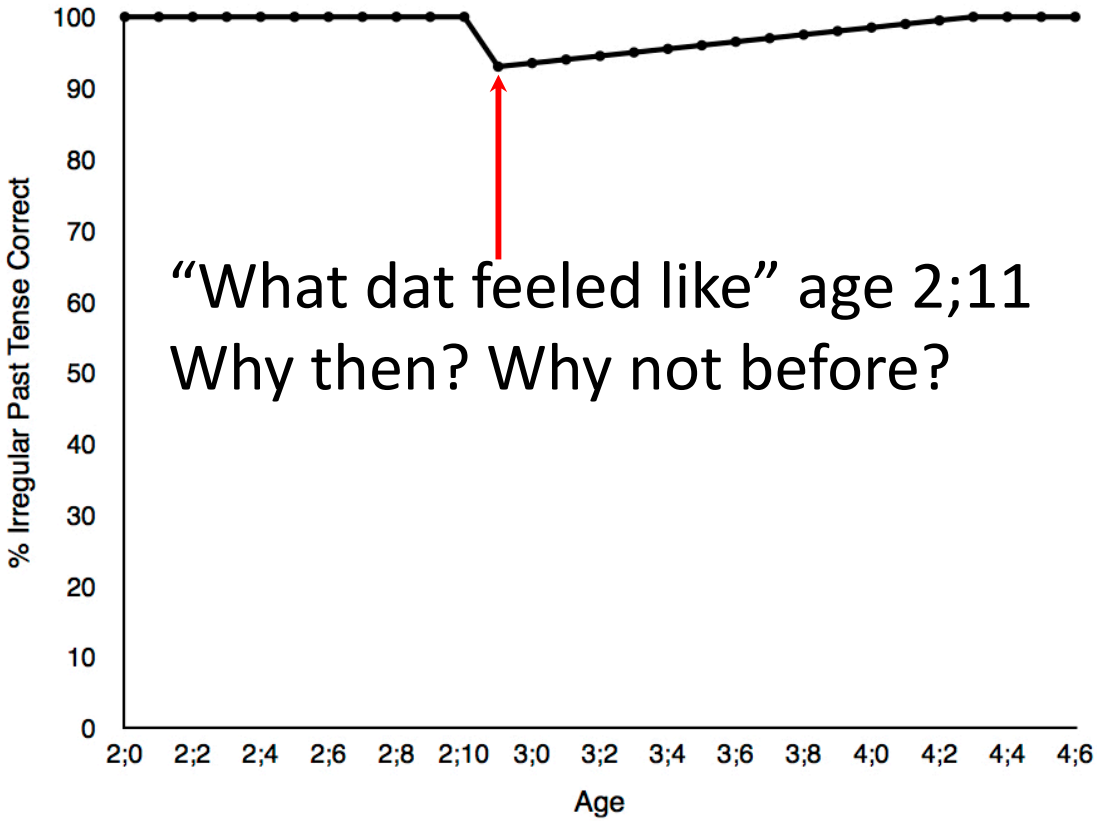
- About 150 exceptions:

  dig-dug; eat-ate; ring-rang; sing-sung (“irregular” verbs)

- What do kids do?
- Classic pattern

The U-shape pattern

Adam



"What dat feeled like" age 2;11

Why then? Why not before?

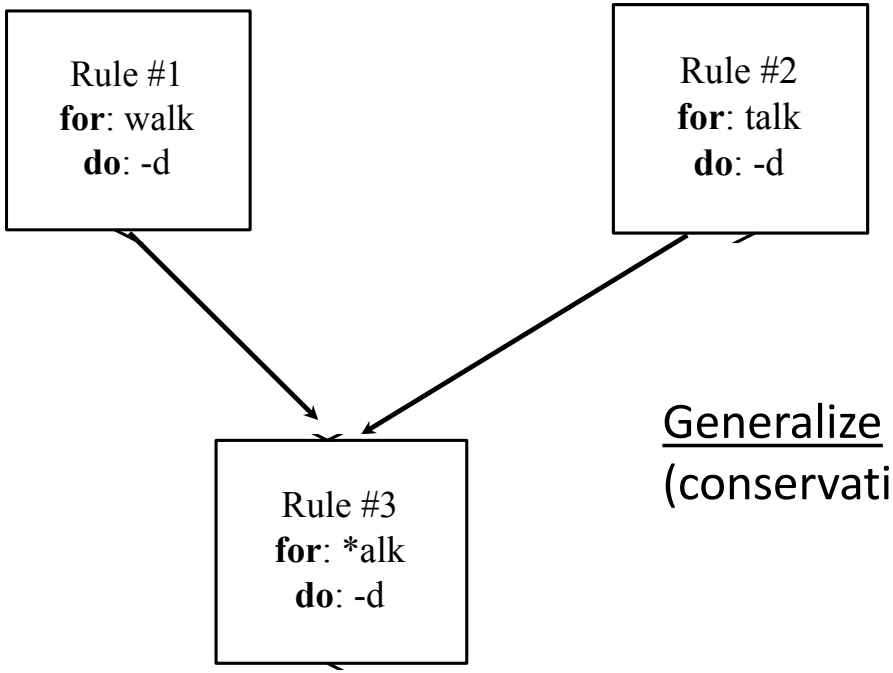
Possible error types for forming past tense of verbs

- go-goed (extend regular rule to an irregular verb → over-regularization)
- bring-brang (extend irregular subgroup to irregular verb, compare sing-sang → over-irregularization)
- (extend irregular to some entirely new form)
- Remember Sussman-Yip? How is the past tense rule formed?

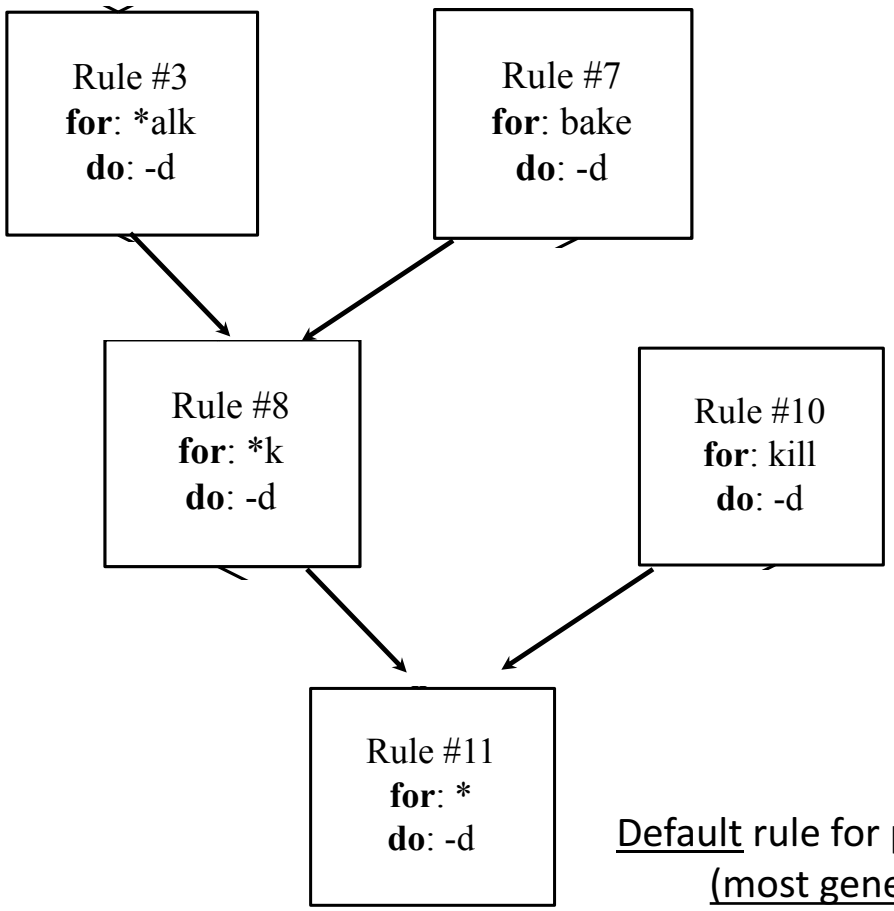
What about learning the English past tense rule? Sussman & Yip

Child hears "…walk…"; … "walked"

Child hears "…talk…"; … "talked"



Generalize by intersecting
(conservatively – why conservative?)



Default rule for past tense – applies everywhere (most general) – might be too general

Regular rule plus a small set of special case "subrules" for some irregular verbs

This actually uses "distinctive features" – the "add d" rule is really this:

→ d / [+voice, −coronal] (close-clos/d/)
→ t / [−voice, +strident] (bake-bake/t/)

1. Internal syllable → ɔt/ (e.g., *think-thought*, *catch-caught*, *buy-bought*)
2. i → æ / N (*sing-sang*, *ring-rang*, ...)
3. i → e / C [d, t] (*feed-fed*, *lead-led*, *meet-met*, ...)
4. [aI, O, o] → u/ [l, n, r] (*fly-flew*, *grow-grew*, *know-knew*, *draw- drew*, ...)
5. Regular rule: Otherwise: apply the d/t rule

But this is not perfect - Where does this kind of learner go wrong?

Over- and under-generalization for Sussman-Yip model

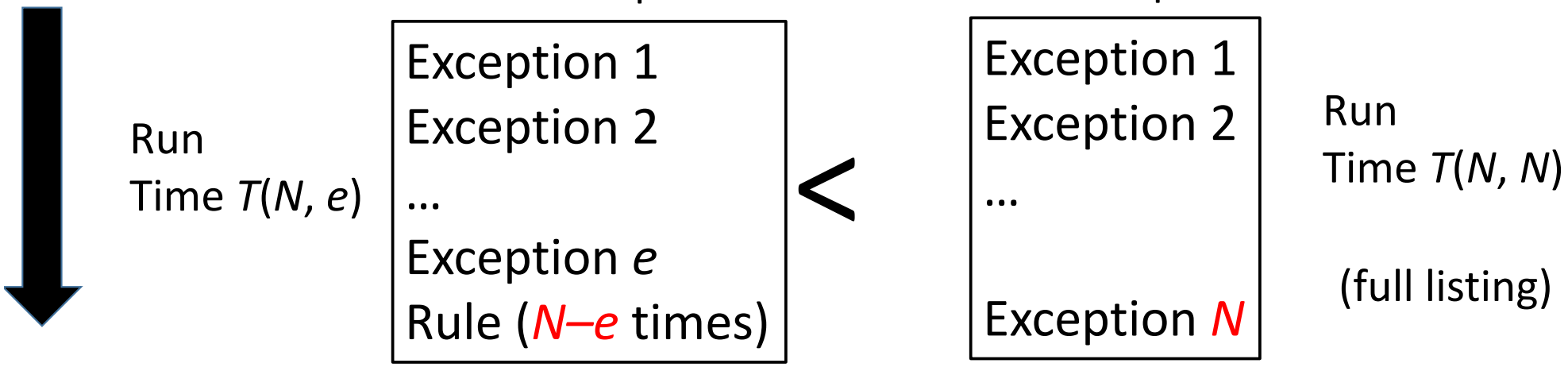
- Overgeneralizes: "special" rule 1 is too general; would convert "love" to "lought" (why else is overgeneral rule possibly bad?)
- Yip-Sussman solves only half the learning problem: Does OK at figuring out the form of rule, but not so good in figuring out when the rules should apply (the scope of rules)

“Elsewhere” principle for rule systems

- How to order the rules? Order such that the most general (least specific) rule is last
- Why?
- Suppose you put most specific rules at the end…what would happen?
- How to order this list of exceptions?
- Order by frequency – keep ordered list of exceptions ranked by frequency, most frequent first
- Not hard to do: “move up” algorithm (Rivest, 1976) – swap just-used item with one just above it (cf. smart phone app list)

Upshot: Kids like it fast

The "tipping point": suppose we want fastest match



Time T to process rule if N items with e exceptions = $T(N, e)$

Time T to process list if N items with N exceptions = $T(N, N)$

Which is faster?

when is $Time(N, e) \leq Time(N, N)$? "Tipping Point"

The Tipping Point: when to use a rule rather than all special case memorized exceptions, given *N* examples

If $Time(N, e) \leq Time(N, N) \rightarrow$ **Using Rule is faster** than listing all examples (a **productive** rule)

The **equality** holds at the tipping point: $\frac{N}{\ln N}$

SO:

- if we are **below** this stipping point, having a rule is faster than just memorizing;
- if we are **above** this point, memorizing full list is faster.

Example: *N= 9*, *N/*ln *N* = 9/2.19=4.096 , so if

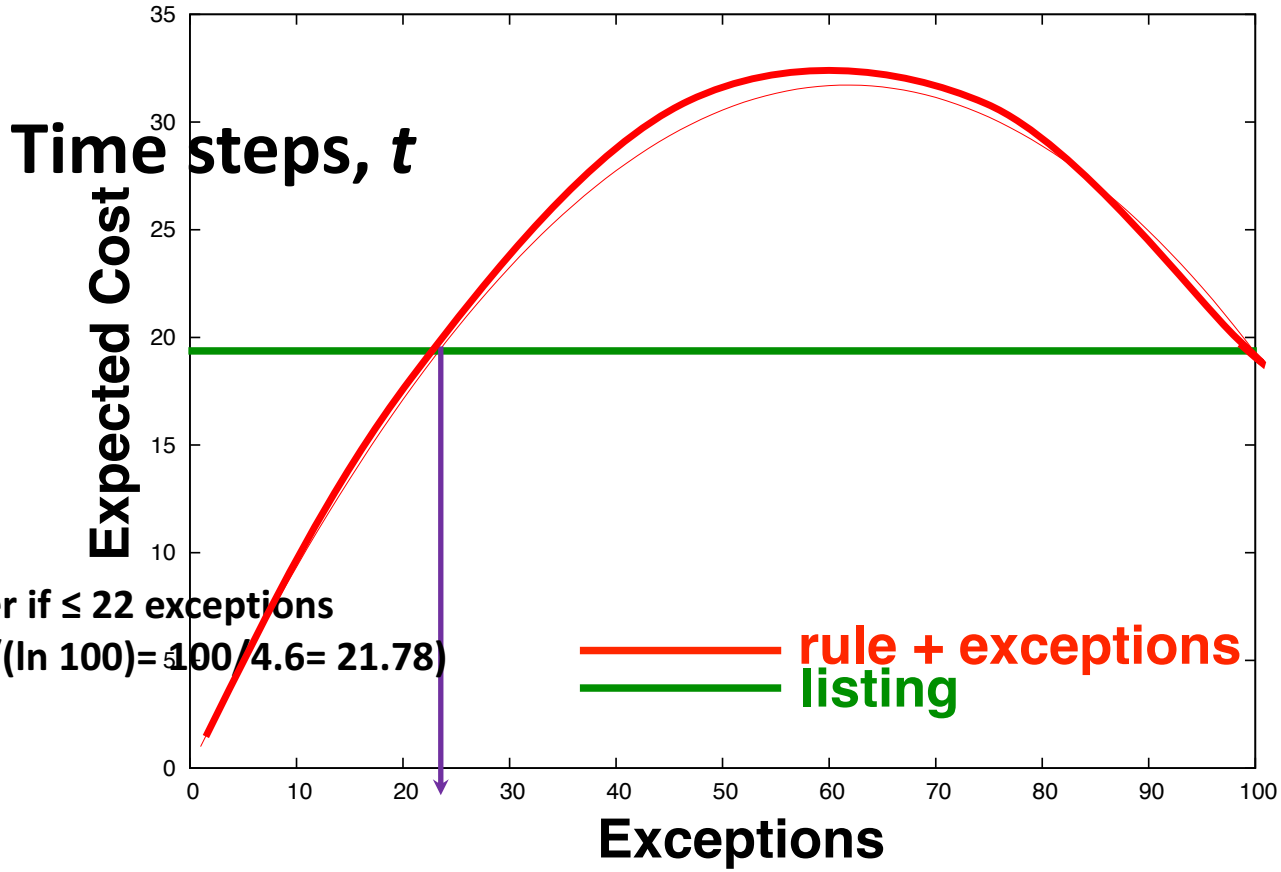
fewer than 4 exceptions → use rule;

if **4 or more exceptions → memorize**

Expected search time

$$e \leq \theta_N := \frac{N}{\ln N}$$

Flat green line= time if all *N* examples memorized as a list (all exceptions)
Red line = time if a rule with *e* exceptions



N=100

Rule is faster if ≤ 22 exceptions
(100/(ln 100)= 100/4.6= 21.78)

Tipping point for # of Exceptions θ given some # of items *N*, θ_N=(N/ln N)

N	θ_N	%
10	4	40.0
20	7	35.0
50	13	26.0
100	23	23.0
200	38	19.0
500	80	16.0
1,000	145	14.5
5,000	587	11.7

Not majority rules!
Smaller # of items – greater fraction of exceptions tolerated before rule judged unproductive, and cheaper to just list all the items – Small data = more rules

And this explains the appearance of rules, and their over-application, etc. in the child language database

Most frequent verbs(total #, # irregular exceptions) bold font = point at which # irregular < tipping pt θ

Top N	*sing→sang*	*feed→fed*	*fly→flew*	*-d*	θ_N
100	—	—	**(8, 3)**	(100, 54)	22
200	**(3, 1)**	—	(11,5)	(200, 76)	37
300	**(3, 1)**	—	(13, 8)	(300, 92)	53
500	**(5, 2)**	**(6, 3)**	(15, 10)	(500, 103)	80
800	(8, 5)	(11, 7)	(18, 13)	(800, 121)	119
1022	(8, 5)	(13, 9)	(22, 16)	**(1022, 127)**	147

In first 100 most frequent verbs, 54 have irregular past tense, e.g., *go-went*, so no chance of rule

But 8 fly-flew type verbs, with 5 examples and 3 exceptions < than t.p. 8/ln 8 = 3.8 → form a rule

blow-blew, *cry-cried*, ***fly-flew***, ***grow-grew, know-knew,*** *lie-lied,* ***throw-threw,*** *try-tried*

Most frequent verbs(total #, # irregular exceptions) bold font = point at which # irregulars < tipping pt θ

Top N	*sing→sang*	*feed→fed*	*fly→flew*	*-d*	θ_N
100	—	—	**(8, 3)**	(100, 54)	22
200	**(3, 1)**	—	(11,5)	(200, 76)	37
300	**(3, 1)**	—	(13, 8)	(300, 92)	53
500	**(5, 2)**	**(6, 3)**	(15, 10)	(500, 103)	80
800	(8, 5)	(11, 7)	(18, 13)	(800, 121)	119
1022	(8, 5)	(13, 9)	(22, 16)	**(1022, 127)**	147

Top 200 verbs: Now 11 fly-flew verbs, with 5 exceptions and θ_{11}= 4, so # exceptions > tipping point 4, so **no rule**

blow-blew, *cry,* ***fly-flew,*** *follow,* ***grow-grew, know-knew,*** *lie, snow, swallow,* ***throw-threw, try***

This generalization would disappear, but…

There are now 3 sing-sang, verbs, with 1 exception, so # exceptions < θ_3 = 2.7

Bring-brought, ***ring-rang, sing-sung***, so the sing-sang rule can persist for a bit

Why Adam "feeled" at age 2 years, 11 months?
"What dat feeled like"

- Entire database: 1022 unique verbs in past tense, 127 are irregular
- $\theta_{1027} = 147$ (so room for more irregulars)

Productivity as function of vocabulary size – top *N* past tense verbs in child corpus
-d column: (# regular verbs, # exceptions found); θ_N = # exceptions tolerated

Top *N*	*sing→sang*	*feed→fed*	*fly→flew*	*-d*	θ_N
100	—	—	**(8, 3)**	(100, 54)	22
200	**(3, 1)**	—	(10, 5)	(200, 76)	37
300	**(3, 1)**	—	(13, 8)	(300, 92)	53
500	**(5, 2)**	**(6, 3)**	(15, 10)	(500, 103)	80
800	(8, 5)	(11, 7)	(18, 13)	(800, 121)	119
1022	(8, 5)	(13, 9)	(22, 16)	**(1022, 127)**	147

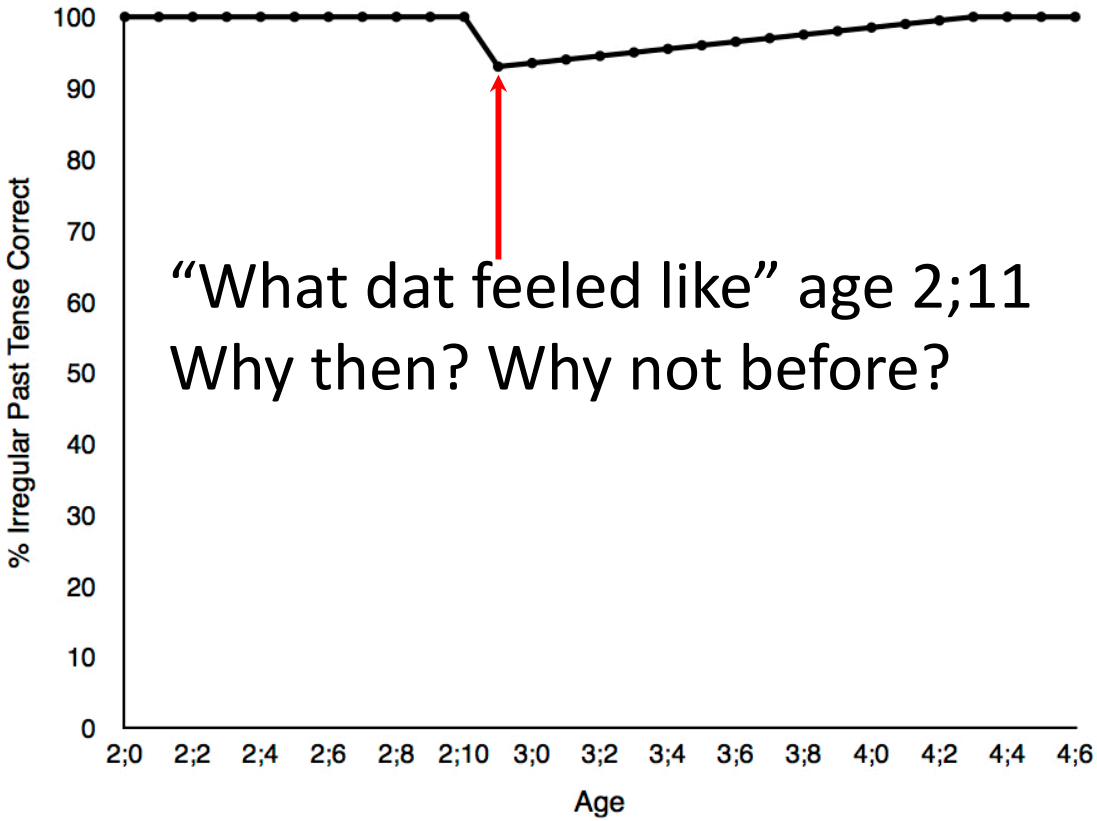
As long as θ_N < # actual exceptions, regular rule is not accepted

Example: *Fly-flew*

1. **blow**, cry, **fly**, **grow**, **know**, lie, **throw,** try (θ_8=3)
2. **blow**, cry, **fly**, follow**, grow**, **know**, lie, snow, swallow, **throw,** try (θ_{11}=4)

The U-shape pattern

Adam



"What dat feeled like" age 2;11

Why then? Why not before?

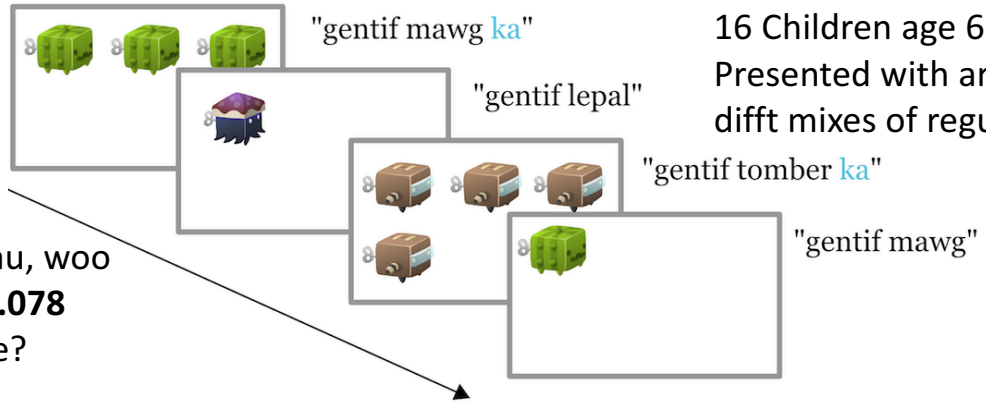
Why Adam feeled at age 2;11 but not before

Tipping point

Top N	*sing→sang*	*feed→fed*	*fly→flew*	*-d*	θ_N
100	—	—	**(8, 3)**	(100, 54)	22
200	**(3, 1)**	—	(10, 5)	(200, 76)	37
300	**(3, 1)**	—	(13, 8)	(300, 92)	53
500	**(5, 2)**	**(6, 3)**	(15, 10)	(500, 103)	80
800	(8, 5)	(11, 7)	(18, 13)	(800, 121)	119
1022	(8, 5)	(13, 9)	(22, 16)	**(1022, 127)**	147

So this is almost <u>exactly</u> when we would expect the regular productive rule to start being applied, and it marks the first appearance of over-application of the rule in Adam's speech

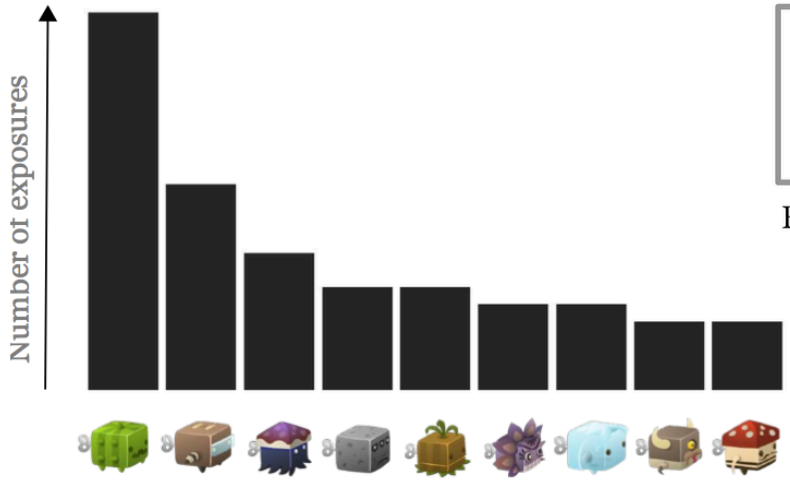
Look at Adam's use of past tense verbs by the age 2;11 – point of emergence of regular rule
By this point, Adam had **300** verbs, of which **57** were **irregular**. θ_{300}=53
So: (nearly) **at tipping point for rule!**



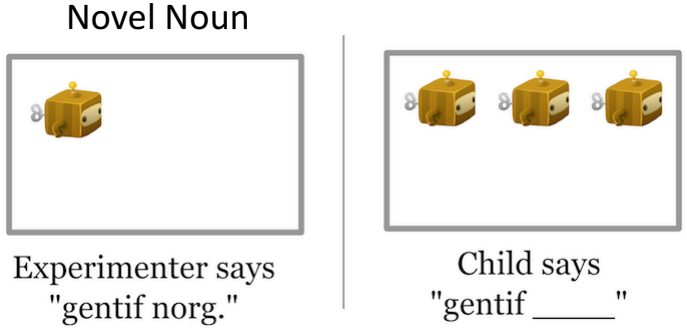
16 Children age 6-8 years old
Presented with artificial ‘sentences’ w/ two difft mixes of regular vs. exceptions to ”plural”

Regular ending: ka
Irregular: po, tay, lee, bae, mau, woo
9 examples, tipping point = 4.078
When will child hold onto rule?

“Nouns” presented according to a power law
Zipf frequency; 27 example sentences;
2 different sequences of regular patterns vs. exceptions



Artificial language “wug” test

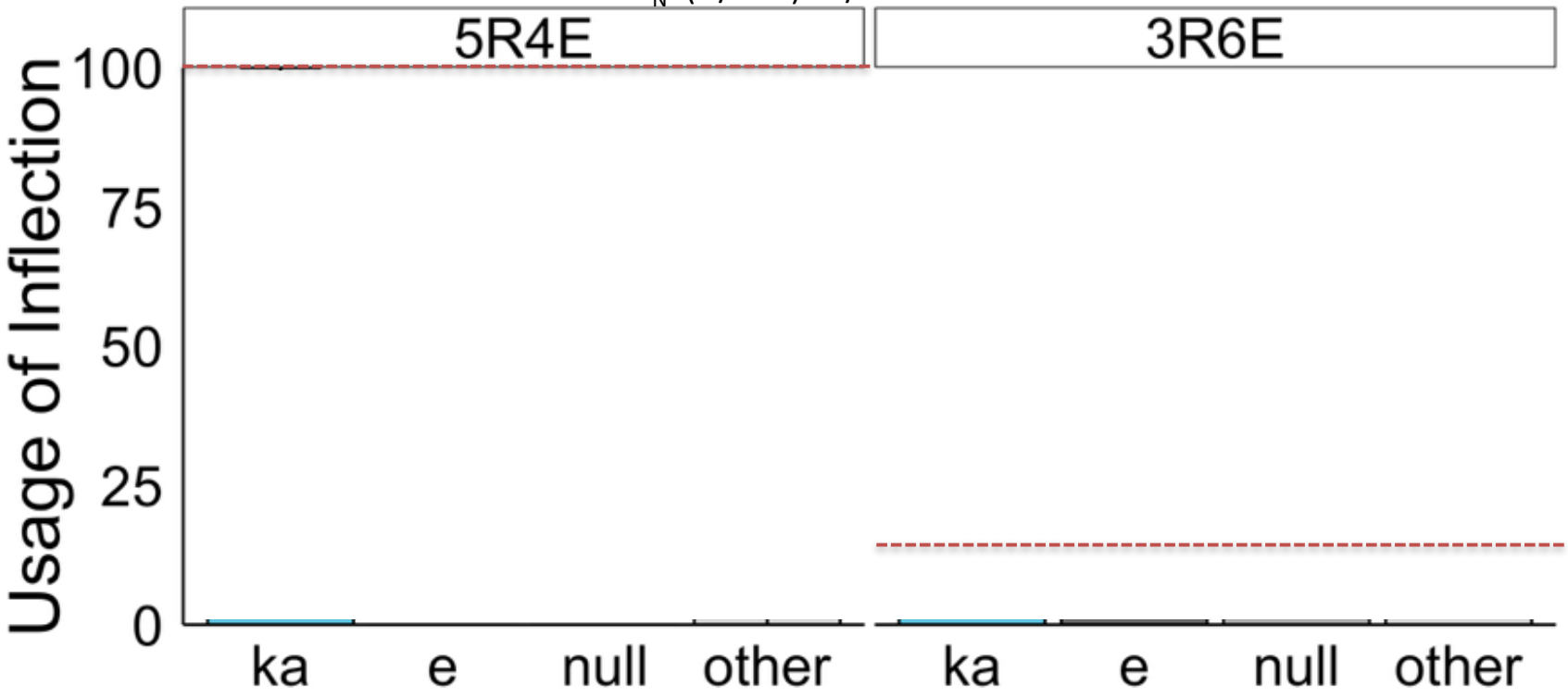


Now test tipping point of 4.078 by creating 2 sets of stimuli, one with regulars > tipping point, one < tipping point: 5, 4 vs. 3, 6

from 16 children age 6-8 years

5 Regular Forms/4 Exceptions (predict: rule) 3 Regular Forms/6 Exceptions (predict: no rule)

θ_N=(N/ln N)= 9/ln 9 = 4.096

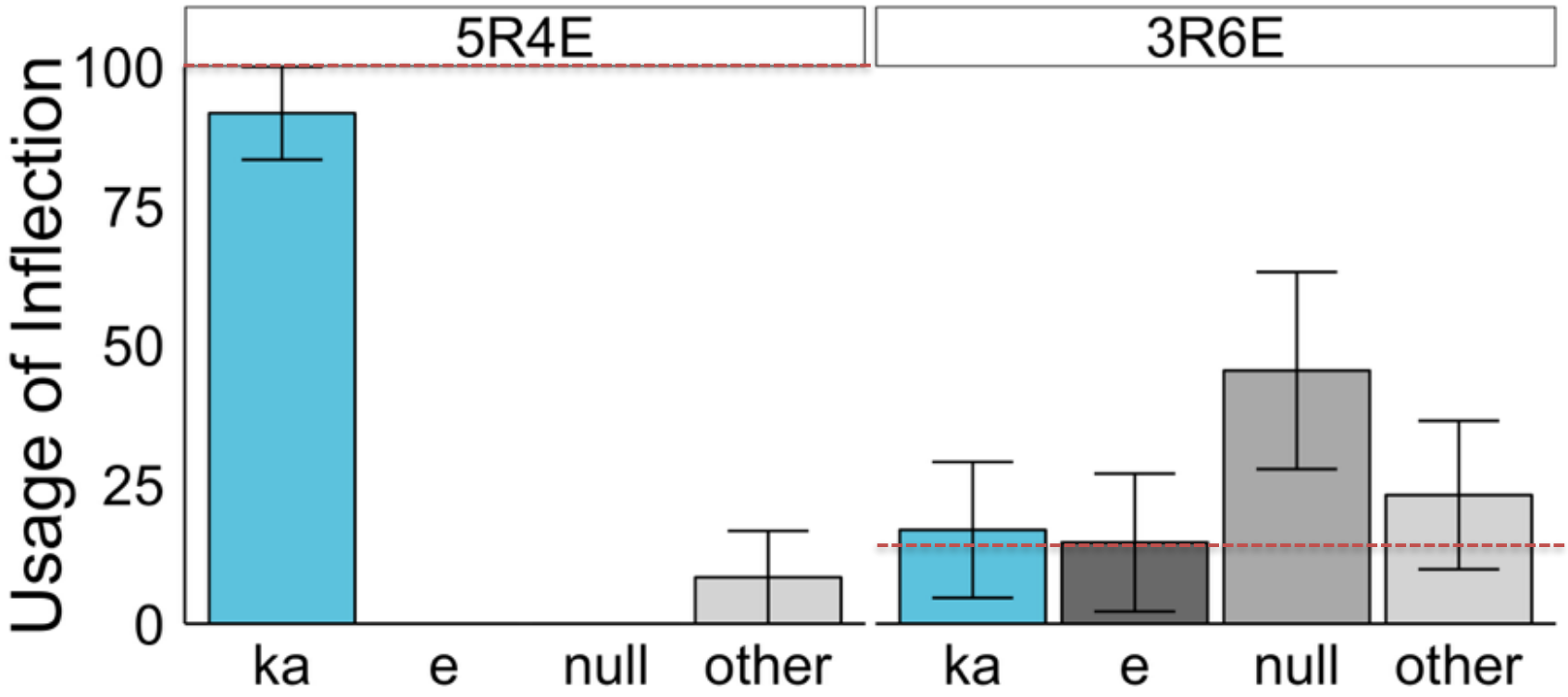


5 regular, 4 not; vs. 3 regular, 6 not (total 9 examples)

from 16 children age 6-8 years

5 Regular Forms/4 Exceptions

3 Regular Forms/6 Exceptions



5 regular, 4 not; vs. 3 regular, 6 not (total 9 examples)

Why less is more: when Small Data succeeds

the ONION
America's Finest News Source

Study Reveals: Babies Are Stupid

Above: Despite their relatively large cranial capacities, babies such as this one are so unintelligent that they are unable to distinguish colorful plastic squeak toys from food sources.