Two mechanisms of language learning:

Rules and statistical learning

1

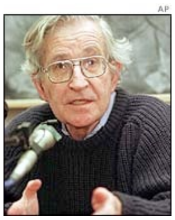
1

Knowledge of language: two questions

- Innateness:
  - What allows a child to learn language *More on this in unit 3…*
- Productivity:
  - What is it that a child learns (today's question)
    - Obviously, the child learns *words*
      - We won't discuss this further…
  - But what is it that allows the child learns, and allows her to generate ***new*** forms (productivity)

2

What is it that a child learns about their language?



- Answer (so far): **Rules**
  - Plural: Noun+S
  - Sentence: NP+VP

3

3

Seriously?

Rules?

4

4

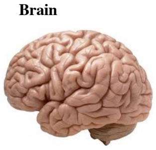
THEY'RE MADE OUT OF MEAT
By Terry Bisson

- "They're made out of meat."
- "Meat?"
- "Meat. They're made out of meat."
- "Meat?"
- "There's no doubt about it. We picked up several from different parts of the planet, took them aboard our recon vessels, and probed them all the way through. They're completely meat."
- "That's impossible. What about the radio signals? The messages to the stars?"
- "They use the radio waves to talk, but the signals don't come from them. The signals come from machines."
- "So who made the machines? That's who we want to contact."
- "They made the machines. That's what I'm trying to tell you. Meat made the machines."
- "That's ridiculous. How can meat make a machine? You're asking me to believe in sentient meat."
- "I'm not asking you, I'm telling you. These creatures are the only sentient race in that sector and they're made out of meat."

5

Rules elicit a mind-body problem

Brain



Mind

Noun+S

NP+VP

How can meat encode NP?

6

6

Today's plan

- Connectionism as an alternative to rules (part 1)
- Productivity has *two* potential sources (part 2):
  - statistical learning
  - rules
- What does it all mean (part 3)

7

Connectionism as an alternative to rules

Part 1

8

The traditional account of inflection

- Regular forms: generated by a rule
  - *Rats, Cats, Dogs*
  - *Liked, cooked..*
- Irregular forms: must be stored in the lexicon
  - *Mice, feet, oxen*
  - *Went, ate…*
- ”regulars” and “irregulars” require different mechanisms
  - Rule
  - lexicon

9

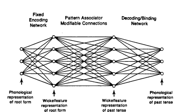
The connectionist alternative

- A single mechanisms forms regulars and irregulars
- No distinction between
  - Categories: *verbs, noun*
  - Instances: *like, dog*
- No rules!

10

Connectionists networks

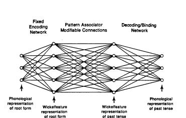
- Capture knowledge as connections between inputs (given) and outputs (outcome)
- Can learn and generalize from training
  - Trained: rat-rats
  - Generalize to: *lat-lats*



11

Rumelhart & McClelland (1986) Past tense model

- What is given during training?
  - Input: Phonological representation of base (e.g., *sit, like)*
  - Output: Phonological representation of past tense (e.g., *sat, liked*)
  - Feedback: right/wrong
- What's learned?
  - Can form correct plurals for existing words
  - Generalize to new forms (e.g., blix)
- How?
  - Compare output to target
  - Adjust *weights* on connections between input and output

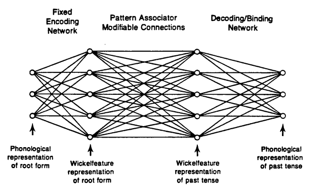


12

Rumelhart & McClelland (1986) Past tense model

Achievements

- Handle both regulars and irregulars on the same network
- Mimic some aspects of language acquisition
- Generalize to new forms
- “brain-inspired”



13

A connectionist alternative

Plurals

	Singular	plural
learn	dog	dogs
	Cat	cats
generalize	dat	dats

How do these models generalize?

- Answer: by similarity
- Simplified account:
  - Train:
    - Dog
    - Cat
  - Generalize: Dat
- How: rely on the overlap (similarity) between training and test items
- Notice: *Brain-inspiration* is rather indirect:
  - These models aren’t about neurons and synapses—what they encode is *information* (cognition) not the brain hardware

14

How does this differ from rules?

15

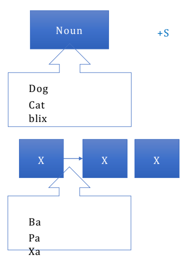
Some examples of rules

- Noun$_{stem}$+S→Noun$_{plural}$
- X→XX

16

Some characteristics of rules

- Operate on entire classes, not specific instances
  - Classes: (e.g., noun, syllable)
  - Instance: (e.g., *dog, ba)*
- Rules do not discriminate:
  - Apply to all members of the class alike, regardless of
    - Familiarity
    - Properties: their sound, meaning….
  - Form *equivalence classes*
- Rules operate on variables (e.g., N)
  - Noun $_{plural}$=Noun $_{singular}$ +S
  - X→XX
  - Blind to instances → generalize across the board



Hypothesis: the mind is algebraic

17

Rules vs. constraints: important! Two meanings of "rules"

In linguistics

- Rules:
  - Algebraic recipes that transform inputs into outputs
- Rules are distinct from *constraints*:
  - Restrictions on outputs
- Note: Constraints are algebraic!
  - E.g., Onset: syllables must have an onset
    - Onset/syllable: equivalence classes
    - Onset defines an algebraic relation between them

Broadly (as I use it here)

- Here: I use "rule" generically, to refer to *all* operations over variables, regardless of whether they apply to
  - Input: ("rules" in linguistics"
  - Outputs ("constraints" in linguistics")

18

Rules and the mind-body problem

Some believe that rules invoke mind-body Dualism

- Rules are ”abstract”
- Abstraction is “ethereal”, aren’t part of the body
- hence: rules are ethereal, cannot be in the brain

The PSSH [physical symbol system hypothesis] makes a Cartesian distinction between thought and action, treating mind as disembodied. That is, according to PSSH, the exact same thoughts occur when a computer is manipulating symbols by using rules and when a person is manipulating the same symbols by using the same rules. The particulars of the body housing the symbol manipulation were thought to be irrelevant.

Glenberg AM, Witt JK, Metcalfe J. From the Revolution to Embodiment: 25 Years of Cognitive Psychology. *Perspectives On Psychological Science: A Journal Of The Association For Psychological Science.* 2013;8(5):573-585.

19

Rules and the mind-body problem

Brain

Mind

Socrates is a man
Every man is mortal

Socrates is mortal

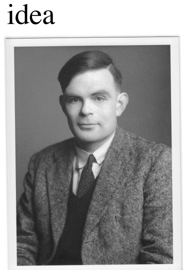
How can meat think?

20

20

Alan’s Turing revolutionary idea

- Thinking is a physical process
- It is physical laws (not some Cartesian stuff) that makes thinking happen

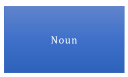


21

Turing's revolutionary idea

Symbols: a two-sided "coin"

- Form (physical)
  - e.g., a square
- Meaning (information)
  - E.g., "noun"



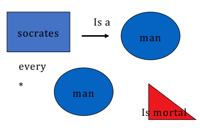
Thinking as a physical process

- Machines can manipulate form (a physical operation)
- By manipulating the physical form, you can systematically manipulate meaning
  - *Note: it is the **physical** structure of symbols that makes thinking happens—not Dualists at all!*
- This can capture thinking
- *The catch: you need to structure your symbols correctly*….

22

Program and symbols

1. find --->
   copy figure to the left of --> to bottom line
2. find figure to the right of -->
   make a cutout of that figure and place in a buffer
3. find figure to the right of *
   copy that figure to buffer
4. compare the contents of the buffer:
   - if
     - match-->continue-->move two figures right of * and copy that figure to bottom line
   - else: stop



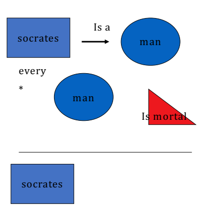
Adapted from Pinker (1994)

23

23

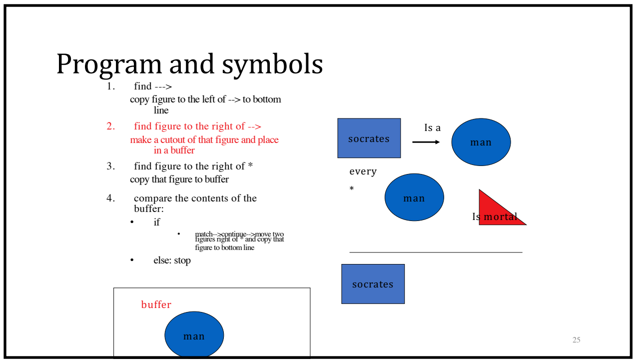
Program and symbols

1. find --->
   copy figure to the left of --> to bottom line
2. find figure to the right of -->
   make a cutout of that figure and place in a buffer
3. find figure to the right of *
   copy that figure to buffer
4. compare the contents of the buffer:
   - if
     - match-->continue-->move two figures right of * and copy that figure to bottom line
   - else: stop

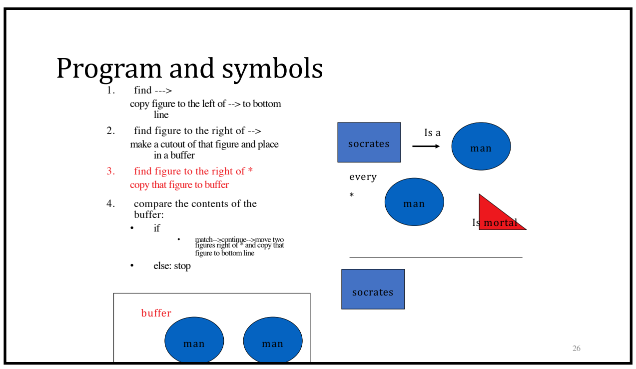


24

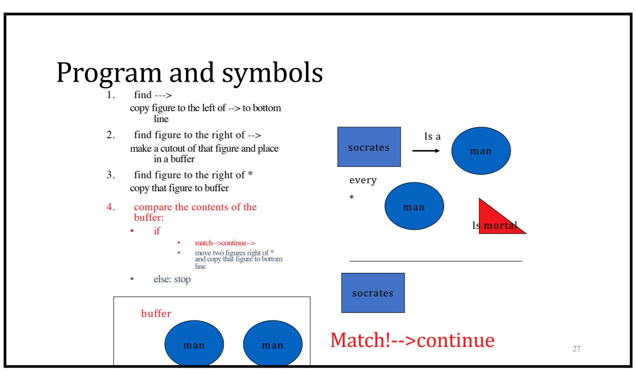
24


Program and symbols
1. find --->
copy figure to the left of --> to bottom line
2. find figure to the right of -->
make a cutout of that figure and place in a buffer
3. find figure to the right of *
copy that figure to buffer
4. compare the contents of the buffer:
• if
• match-->continue-->move two figures right of * and copy that figure to bottom line
• else: stop
socrates
Is a
man
every
*
man
Is mortal
socrates
buffer
man
25

25


Program and symbols
1. find --->
copy figure to the left of --> to bottom line
2. find figure to the right of -->
make a cutout of that figure and place in a buffer
3. find figure to the right of *
copy that figure to buffer
4. compare the contents of the buffer:
• if
• match-->continue-->move two figures right of * and copy that figure to bottom line
• else: stop
socrates
Is a
man
every
*
man
Is mortal
socrates
buffer
man
man
26

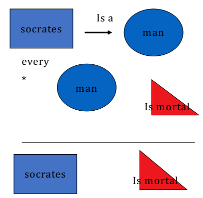
26


Program and symbols
1. find --->
copy figure to the left of --> to bottom line
2. find figure to the right of -->
make a cutout of that figure and place in a buffer
3. find figure to the right of *
copy that figure to buffer
4. compare the contents of the buffer:
• if
• match-->continue-->
• move two figures right of * and copy that figure to bottom line
• else: stop
socrates
Is a
man
every
*
man
Is mortal
socrates
buffer
man
man
Match!-->continue
27

27

Program and symbols

1. find --->
   copy figure to the left of --> to bottom line
2. find figure to the right of -->
   make a cutout of that figure and place in a buffer
3. find figure to the right of *
   copy that figure to buffer
4. compare the contents of the buffer:
   - if
     - match-->continue-->
     - move two figures right of * and copy that figure to bottom line
   - else: stop



28

28

What makes the program work? The computational theory of mind (Fodor)

Representations

- Information is encoded by structured representation
  - Form 

  - Meaning: Socrates; *dog+s*
- Form and meaning are systematically linked
  - Atomic meaning (dog) gets atomic form
  - Complex meaning (*dogs*) gest complex form

Operations

- Structure *causes computation to happen*
  - Each step only depends on form (e.g., square), not meaning…

Fodor, J., and Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. Cognition 28, 3-71.

29

29

Consequence

- Systematicity
  - If you know something about John and Bill
    - *John and Bill are nice*
    - then you know something about Bill
      - *Bill is nice*
  - Note: systematicity is not merely possible; it's **inevitable…**
- Productivity: generalization across the board
  - Even when a new item is utterly dissimilar to test items…

30

Eliminative connectionism rejects these assumptions

Connectionist representations: associations

- Mental processes depend on association, not constituent structure
- *Note:* this is not necessarily the case for *all* forms of connectionism, but it *is* likely the case in the popular networks that are on the market….
  - *More soon…*

CTM: Structure sensitive operations

- Mental processes are caused by structure

31

What's at stake

- Connectionism has challenged Chomsky's research program
  - No such thing as rule
  - No universal grammar either…

32

The rejection of rules

Characterization of performance as **'rule-governed' are viewed as approximate descriptions** of patterns of language use; no actual rules operate in the processing of language.

McClelland JL & Patterson K (2002) Rules or connections in past-tense inflections: what does the evidence rule out? *Trends Cogn Sci 6(11):465-472*

33

Agenda

- What do people do?
  - Do they represent rules
  - Are rules "the only game in town" or do people also track statistical association (like connectionist networks)?
- **Spoiler alert**: yes, people use both…
- How to tell which one plays a role in a specific case?

34

Two mechanisms of learning: statistical learning and rules

Part 2

35

35

How many words?

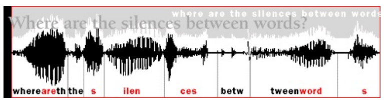
- *Prettybaby?*
- *Wannahelpme*
- A foreign language: can you tell how many words

36

36

The problem of speech segmentation

- Problem: speech is continuous
  - No boundaries between words
- Challenge: how can infants ever acquire words if they cannot segment them?
  - How do they discover what is a word?
- Answer: by statistical learning!



37

37

Are you a good statistical learner?

- Listen to the speech stream

38

38

Have you heard this "word"?

1. ?
2. ?
3. ?
4. ?
5. ?
6. ?
7. ?
8. ?
9. ?
10. ?
11. ?
12. ?
13. ?
14. ?
15. ?
16. ?

39

39

Have you heard this "word"?

1	bulado	W
2	ladobi	N
3	tibata	N
4	dobigo	N
5	bigoku	W
6	datiba	W
7	dupabu	N
8	tadupa	W
9	tibata	N

10	dobigo	N
11	dupabu	N
12	bigoku	W
13	bulado	W
14	ladobi	N
15	datiba	W
16	tadupa	W

40

40

What is going on?

41

41

What's going on?

Phase 1: familiarization

- Four repeated "words"
  - **bigoku**
  - **bulado**
  - **datiba**
  - **Tadupa**
- Random permutations:

bigokutadupadatibatadupadatibabulado tadupadatibabigokutadupabuladobigoku bigokubuladodatiba

Phase 2: test

- compare
  - Words: **bigoku**
  - Nonwords:
    - **ladobi**
- Finding: words stand out. Why?
- Answer: people track *statistical information*

42

42

How statistical information helps?

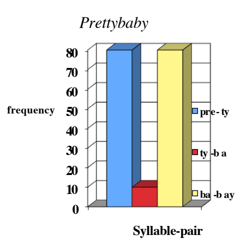
- Transitional probability: the statistical probability that a certain event (Y) can occur given the probability of another event (X)

  Y|X=frequency of XY/Frequency of X

  - *Example: Prettybaby*
    - High probability within word: *Pre-tty*
    - Low probability across words: *Ty-bay*
- Statistical probability can tell us whether certain sound combinations form a word!

43

Using statistical troughs in word segmentation



- Word boundaries are marked by *troughs* in transitional probability
- Statistical information can help segment words

44

Statistical learning can help discover words

- Iseetheprettybabyintheroom
- ababyisstandingneartheprettygirl
- motherfedherbaby

45

Statistical learning can help discover words

- Iseethe prettybabyintheroom
- ababyisstandingnearthe prettygirl
- motherfedherbaby

46

Are infants sensitive to statistical structure?

Saffran Aslin, Newport *Science*, *274* (1996)

Also: sections 4.0-4.2 in textbook

47

Approach: artificial language

- Artificial language: an invented “language”, constructed according to some specific regularities
  - Advantages
  - disadvantages

48

Experiment 1

- **Familiarize:**
- **Tupirogolabubidakupadotigolabutupiropadotibidaku**

tupiro
golabu
bidaku
padoti

- **Test**: a single item repeated (e.g., **tupiro tupiro tupiro tupiro),**

tupiro	word	(after every *tu* there is *pi*)
golabu	word	(after every *go* there is a *la*)
dapiku	NW	(*da* is never followed by *pi*)
tilado	NW	(*ti* is never followed by *do*)

49

49

Full design

- **Familiarize: tupirogolabubidakupadoti (no breaks)**

condition A	condition B
tupiro	dapiku
golabu	tilado
bidaku	burobi
padoti	pagotu

- test

tupiro	word	NW
golabu	word	NW
dapiku	NW	word
tilado	MW	word

50

50

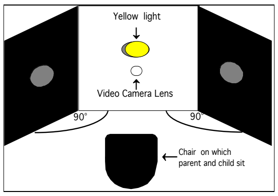
Preferential looking time

https://video.search.yahoo.com/yhs/search;_ylt=AwrCxGGykI9cdxoAE.8PxQt.;_ylu=X3oDMTByMjB0aG5zBGNvbG8DYmYxBHBvcwMxBHZ0aWQDBHNlYwNzYw--?p=preferential+looking+time+site%3Ayoutube.com&fr=yhs-sz-001&hspart=sz&hsimp=yhs-001

51

51

Testing procedure
a. fixation

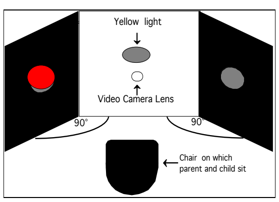


52

52

After fixation is verified

Bapaga…



Rationale:
•Novel →look longer

53

53

Finding:

Listening times (sec)

Familiar	novel
7.97	8.85

54

54

Experiment 2:

- Discriminate words from part-words
- A real word illustration
  - **Words**: Pretty, baby
  - Part words: tebay
- Saffran et al used artificial words

55

55

Experiment 2

- familiarize:
  Tibudopabikudaropigolatupabikutibudo
- test: compare
  - Words: **pabiku pabiku pabiku**
  For words: (p=1)
  - Part words: combinations of word parts
    pigola part word (p=.33)
    tudaro part word (p=.33)

56

56

Finding:

- longer listening time for nonwords

Listening times (sec)

Familiar	novel
6.77	7.60

57

57

Conclusion:

- Infants readily extract statistical information
- Use statistical information to segment words

58

58

In an experiment, people heard the "words" *baga, dama, topo* in random order. Where would you expect a *frequency trough*?

A. balga
B. Malto
C. Tolpo
D. Dalma

59

In an experiment, people heard the "words" *baga, dama, topo* in random order. What result would you expect (looking time)?

A. Baga>dama
B. Mato>baga
C. Baga>mato

60

Why does it matter?

- Helps segment speech
  - Necessary to learn words
- *An alternative to rules!*
- *Also: critical in modern AI*

61

61

How do infants acquire their language?

- Answer (so far): by relying on rules?
- Answer (now): statistical learning matters too!

62

62

Questions

- What's the difference between rules and statistical learning
- Do we need both?
  - Is statistical learning sufficient to capture language?

63

63

Outline

- Beyond rules: the role of statistical associations
  - People track statistical information
- Why rules are also necessary
  - Rules vs. statistical associations: what's the difference
  - Infants also learn rules
- Conclusion: two sources of productivity
  - Rules
  - Statistical learning (associations)

64

64

Learning rules

65

65

Listen to these words in a new language

66

66

Which of these words is likely to come from that language?

- *Wofewo*
- *wofefe*

67

67

questions

- How did you tell words from nonwords?
- How does this type of learning differ from statistical learning?

68

68

Word list

Words
gatiti
ganini
galala
linana
lititi
ligigi
nigigi
nititi
nilala
talala
tatiti
tanana

Filler Words
gatiaga
ganaga
gagiga
galaga

Test words:

- Wofefe ("word")
- Wofewo (nonword)

Question: can statistical learning(the co-occurrence of specific syllables) help figure it out?

69

69

Word list

Words
gatiti
ganini
galala
linana
lititi
ligigi
nigigi
nititi
nilala
talala
tatiti
tanana

Filler Words
gatiaga
ganaga
gagiga
galaga

Test words:
- Wofefe ("word")
- Wofewo (nonword)

- No!
  - Transitional probability:
    - Wofefe=0
    - Wofewo=0
- How do we learn these words, then?

70

70

Finding structure

- Words follow abstract structure: ABB
  - Note: A, B stand for *any syllable*
    - Abstract categories
    - Not specific syllables
- Word/nonword contrast defined by structure alone
  - *Wofefe* =ABB
  - *Wofewo=ABA*
- Structure can define "words" even when statistical learning cannot!

Words	structure
gatiti	ABB
ganini	ABB
galala	ABB
linana	ABB
lititi	ABB
ligigi	ABB
nigigi	ABB
nititi	ABB
nilala	ABB
talala	ABB
tatiti	ABB
tanana	ABB

71

71

How do we discover the structure here?

- ~~Statistical information~~
- Rules!

72

72

What is a rule?

- Rules operate on entire classes (e.g., Noun) using variables
  - $\text{Noun}_{singular}$ + s → Noun_{plural}
  - X→XX
- Consequence: generalization across the board to *any* novel instance

Class	*instances*
verb	*like, think, see, grop*
Noun	*dog, cat, blix*

73

73

Examples of rules

example		Rule
English Plurals	Bat-bats cat-cats	Noun +S
Reduplication (Ilocano, Philippines):	Pusa **pus**pusa (cat - cats) Kalding **kal**kalding (goat-goats)	XX

74

74

Rule vs. statistical learning: what's the difference?

- Critical difference: classes vs. instances of a class

Class (variable)	Instances
Noun	*Dog, cat, mouse*
Verb	*Sit, talk, walk*
A ("Any syllable")	*Ba, ga, la*

75

75

Rule vs. statistical learning: what's the difference?

What is a rule?
- Rules typically operate on ***entire classes***
  - Nouns, not *dog*
- Operate on *variables*
  - $Noun_{singular}$ + s → $Noun_{plural}$
  - XX→X+X

Statistical learning
- Tracks the co-occurrence of ***specific elements***
  - *Bi-go-ku*
  - *Ta-li-ru*
- Note:
  - no abstract classes
  - No variables

76

76

Rule vs. statistical learning how they get the job done?

	Rules	Statistical learning
What the track?	Structure: relation between variables (e.g., Noun+S)	Transitional probability of specific instances (e.g., dog+s)
What is a "word"?	If a form has the right structure → word	If word is familiar (frequency is high) → word
Generalize	By operating on variables	Associating specific instances

77

77

Can infants learn rules?

Marcus et al., *Science*, *(1999)*

78

78

Experiment 1

- Familiarize : group 1
  ABA
  ga ti ga
  li na li
  la ta la
- Test:
  - Consistent:
    ABA
    wo fe wo
  - Inconsistent:
    ABB
    wo fe fe

79

79

Finding:

Listening times (sec)

Familiar	novel
6.3	9.0

80

80

what did the babies learn?

- rule ABA/ABB
- statistical information: voicing
  - voice: vibration of vocal chords
    - e.b, g=voiced
    - p, t=unvoiced
- Statistical correlate of ABA in Exp. 1:
  - voiced-unvoiced-voiced
  - Familiarization: ga-ti-gu
  - Test: wo-fe-wo

81

81

Experiment 2

- familiarize:

	group 1	group 2
	ABA	ABB
	ledile	ledidi
	lejele	lejeje

- Test:
  - Consistent:

	ABA	ABB
	bapoba	bapopo

  - Inconsistent:

	ABB	ABA
	bapopo	bapoba

82

82

Findings

Listening times (sec)

Familiar	novel
5.6	7.35

83

83

Experiment 3

- familiarize: group 1 group 2

	AAB	ABB
	leledi	ledidi
	leleje	lejeje

- Test:
  - Consistent:

	AAB	ABB
	babapo	bapopo

  - Inconsistent:

	ABB	AAB
	bapopo	bapbao

84

84

finding:

- longer listening time for inconsistent sentences

Listening times (sec)	
Familiar	novel
6.4	8.5

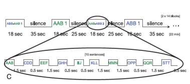
85

85

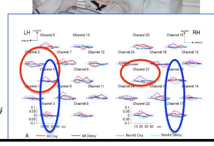
Binding at Birth: The Newborn Brain Detects Identity Relations and Sequential Position in Speech

Judit Gervain[1,2], Iris Berent[3], and Janet F. Werker[4]

- ABB: *bogugu*, *motiti*, etc.
- ABB/AAB *bogugu*, momoti…



Journal of Cognitive Neuroscience 24:3, pp. 564–574



86

Rules vs. statistical learning

	Statistical learning	Rules
What people track?	Co-occurrence of instances Go-no-bu Lo-bo-ga	Relation among variables XXY
What defines a "word"?	frequency	structure
Generalization possible?	Yes	yes
How?	based on association e.g., hearing go-*no*	based on structure
Does similarity to familiar instances matter?	yes	No

87

87

exercise

- An infant hears the following stimuli
  - Po-ga
  - Mi-to
  - Si-no
- Test: compare looking time for:
  - Po-ga
  - Ba-za

88

88

Hear: Po-ga; Mi-to; Si-no
Test: Po-ga vs. Ba-za
Expected looking time results?

A. Baza>poga
B. Poga>baza
C. poga=baza

89

What mechanism?

A. Statistical learning
B. Rule learning
C. Cannot tell

90

Task: hear: *gagada, babana*
test: *momoko* vs. *komomo*
Expected result (looking time)?

A. Momoko=komomo
B. Momoko>komomo
C. Momoko<komomo

91

What mechanism?

A. Statistical learning
B. Rule

92

Hear: bagaga, balala, badada
test: bamama vs. mamafa
What mechanism (think carefully)?

A. rule.
B. Statistical learning
C. Cannot tell

93

Rule or statistical learning, how to tell?

- What is the pattern?
  - Specific “words” (e.g., ba pa) co-occur → statistical learning
  - Abstract structure present (ABA, ABB)→ask:
    - Is there *also* relevant statistical information (e.g., all familiarization items begin with same sound)
      - Statistical pattern is found →*either* rule *or* statistical information can be responsible: cannot tell which one…
      - No statistical information is found →rule is used

94

94

What does this all mean?

Part 3

95

But what does it *really* mean?

- Suppose you find that both rules and statistical learning play a role
- What does it mean about the adequacy of popular connectionist networks?
  - Are they a plausible model of language/cognition?
- What does it mean about *current* deep learning techniques?

96

The rejection of rules: what do people *really* mean?

Characterization of performance as **'rule-governed' are viewed as approximate descriptions** of patterns of language use; no actual rules operate in the processing of language.

McClelland JL & Patterson K (2002) Rules or connections in past-tense inflections: what does the evidence rule out? *Trends Cogn Sci 6(11):465-472*

This is ambiguous between two contradictory claims

1. Rules do not exist in the mind
   - statistical learning is the only game in town,
   - connectionist networks can do it…
2. Rules exist in the mind
   - Although connectionist networks start with no rules, rules can ultimately "emerge"

97

Spelling it out

- Suppose you start with a typical connectionist network
  - Only connections between specific instances and their components
    - e.g., between letters
- Will a rule emerge?

98

How do you tell there is a rule? Two hallmarks of rules

- Systematicity: if you know
    - *Big cup*
    - *Red up*
  - You know
    - *Big red cup*
  - *connectionist networks don't (Lake & Baroni)!*
- Across the board generalizations (Marcus)

1. Lake, B.M., and Baroni, M. (2017). Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In Retrieved from arXiv:1711.00350

99

Across-the-board generalizations

Rules generalize across the board

- Rules generalize to any novel instance, irrespective of whether its features
  - have been all included in the training items (test items fall *within* the training space)
  - Its features have not been trained on (test item falls *outside* the training space)
- Formally: rules generalize beyond the training space

So do people…

- People generalize the identity function to any novel instance
  - Even when feature values are unattested
  - *Kathath* vs. *thathak*
- Even across modalities: from speech to sign….

What about connectionism?

- Berent, I., Marcus, G. F., Shimron, J., & Gafos, A. I. (2002). The scope of linguistic generalizations: evidence from Hebrew word formation. *Cognition, 83*(2), 113-139.
- Berent, I., Bat-El, O., & Vaknin-Nusbaum, V. (2017). The double identity of doubling: Evidence for the phonology-morphology split. *Cognition, 161*, 117-128. doi:10.1016/j.cognition.2017.01.011

100

A simple example

	Input			output			
train	d	o	g	D	O	G	s
	c	a	t	C	A	T	s
Test 1 (within the training space)	d	a	t	D	A	T	s
Test 2 (outside training space)	B	A	T	B	A	T	s

- Connectionist networks generalize within the training space
- Fail to generalize beyond the training space

101

Testing generalizations outside the training space

Logic

- Train on a network on the identity function
  - X—>X
- Hold one feature constant (last=odd)
- Test generalization to untrained feature
  - (last=even)
  - Note: this dissociates identity from similarity
- Finding: the network generalizes according to similarity, not identity

TABLE 1
A Sample Function

Training cases

Input					Output				
0	0	0	1	0	0	0	0	1	0
0	0	1	0	0	0	0	1	0	0
0	0	1	1	0	0	0	1	1	0
0	1	0	0	0	0	1	0	0	0
0	1	0	1	0	0	1	0	1	0
0	1	1	0	0	0	1	1	0	0
0	1	1	1	0	0	1	1	1	0
1	0	0	0	0	1	0	0	0	0
1	0	0	1	0	1	0	0	1	0
1	0	1	0	0	1	0	1	0	0
1	0	1	1	0	1	0	1	1	0
1	1	0	0	0	1	1	0	0	0
1	1	0	1	0	1	1	0	1	0
1	1	1	0	0	1	1	1	0	0
1	1	1	1	0	1	1	1	1	0

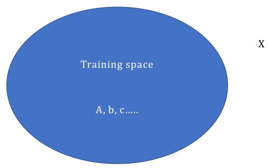
	Input	output
Test items	1 1 1 1 1	1 1 1 1 0

Marcus, G., F. (1998). Rethinking eliminative connectionism. Cognitive Psychology, 37(3), 243-282.

102

The scope of generalization in popular connectionist networks

- Successful generalization *within* the training space (interpolation)
- Failure to generalize *outside* the training space (extrapolation)
- Note: whether any particular item falls within/outside the training space depends on
  - Grain size (e.g., syllable, segment feature)
  - Feature inventory

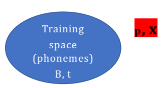


103

The role of representation grainsize

Representing English phonemes

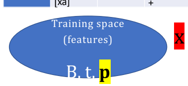
		b	t	p	x
Train	[ba]	+			
	[ta]		+		
Test	[pa]			0	
	[xa]				0



Prediction: no generalization to p, or x

Representing English features

		Labial	Voicing	Velar	Fricative
Train	[ba]	+			
	[ta]		+		
Test	[pa]	+	+		
	[xa]		+	0	0



- By encoding features (rather than segments), we can now fit p (but not x) into the training space

104

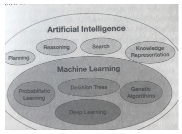
Conclusions

- Connectionist networks fail to generalize beyond the training space
- What did the system learn?
  - Not rules:
    - For any X input, activate X in output
    - Irrespective of training history of
  - Rather: association
    - Output depends on similarity of input to training items
    - Training for each node is independent of others
      - Intuitively: "Unit X does what's likely for unit X"
- Conclusion: a system that lacks rules in the first place does not follow rules spontaneously
  - Rules do not "emerge" in models

105

Deep learning vs. AI

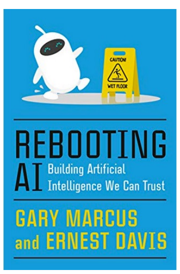
- Deep learning is one form of AI
  - Revolutionizes modern technology
  - Based on "big data"
- Currently, most models that use deep learning rely on associations *only*—no rules!
- Can they "get" language?



106

106

Not really…



107

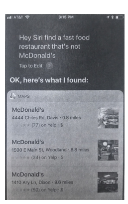
107

Some challenges to "deep learning"

Ask Google…

- Where did Harry Potter meet Hermione Granger?
  - *No answer* (does not name the meeting place)
- What were the seven Horcruxes in Harry Potter?
  - *No answer (no book discusses them as a single list)*
- Problem: cannot integrate information

Ask Siri…



Doesn't get "not"

108

108

Deep learning isn't that deep. Why?

- Only tracks association
- No rules
  - *Unlike people!*

109

109

A hidden innateness problem

- What must be innate in learning mechanism?
  - Connectionism:
    - Representation of instances
    - Learning algorithm
  - Algebraic approach (the computational theory of mind)
    - The capacity to represent variable and operate variables
- Note: two aspects of innateness
  - Properties of learning mechanisms (our question here)
  - Contents of what is learned (the UG question, **not** our issue here)

110

So why is connectionism so popular?

- The strengths of connectionism
  - Many areas do not require generalizations beyond the training space
  - Connectionism can get (that) job done!
- Our weakness of intuitive cognition
  - People confuse the notion of abstraction with Dualism
  - Rules imply some "innate" commitments
    - Some people find this problematic

111

Summary

- People (including infants) can rapidly extract linguistic regularities in linguistic inputs
- Two mechanisms:
  - Statistical learning
  - Rule
- These two mechanisms aren't the same!
- Getting both mechanisms might be critical for the future of cognition AI
- Some of the allure of connectionism arises from our intuitive cognitive biases
- To advance AI and get a better grasp of cognition, we better keep our intuitive biases at bay!

112

112