More past tense!

LINGUIST611

Spring 2022

Administrata

Lab 1 debrief:

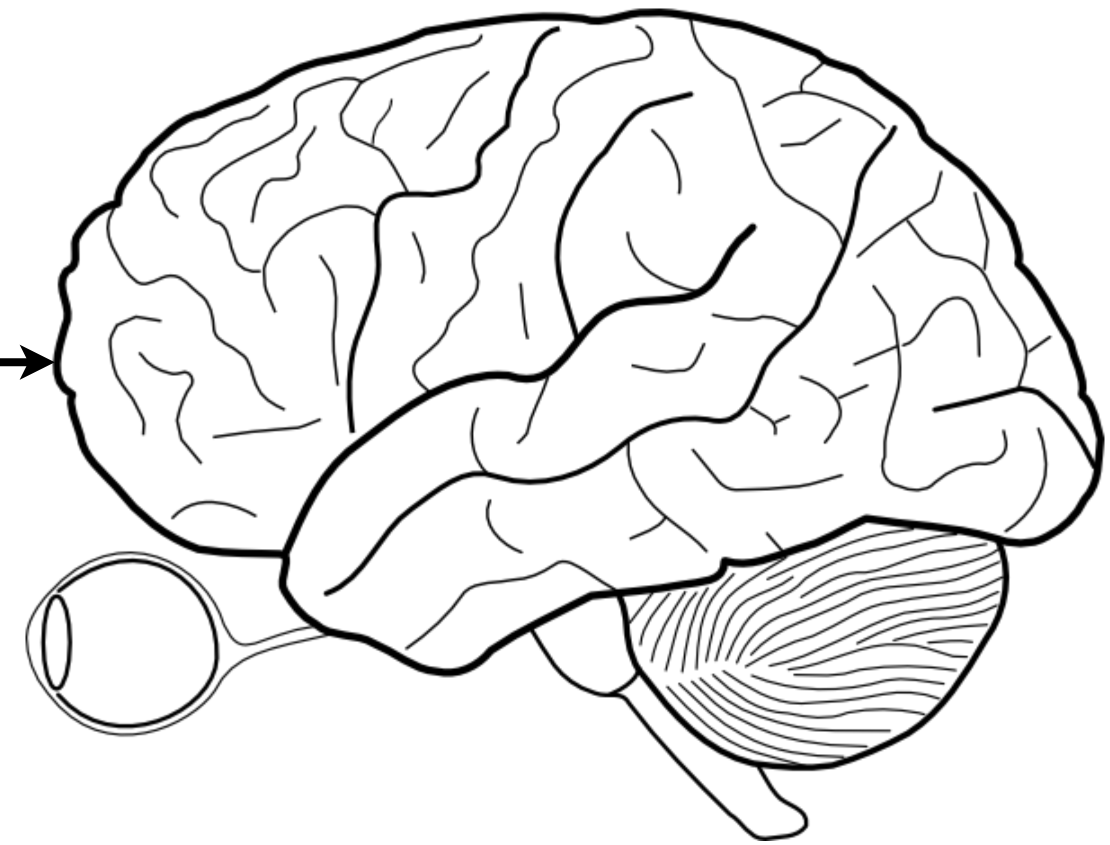
- How well-behaved were your data? Did you replicate the categorical perception effect in English? In Russian? Did you see any interesting patterns in the reaction times? What about the pattern of results in the discrimination task?

- Thoughts on coding up lab reports in R? Was there any major hiccups in coding? Anything you wished you knew how to do?

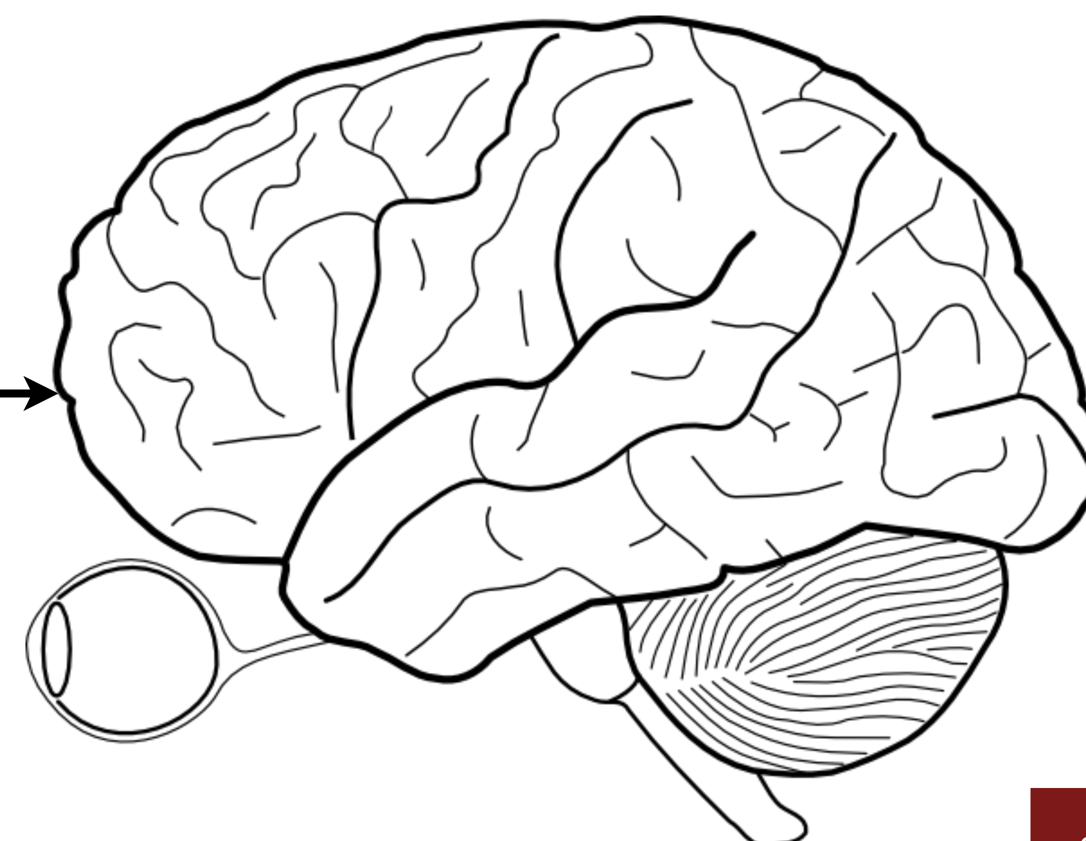
One system or two?

loop - [lup]
please - [pliz]
grade - [gred]
feel - [fil] (PAST: [felt])

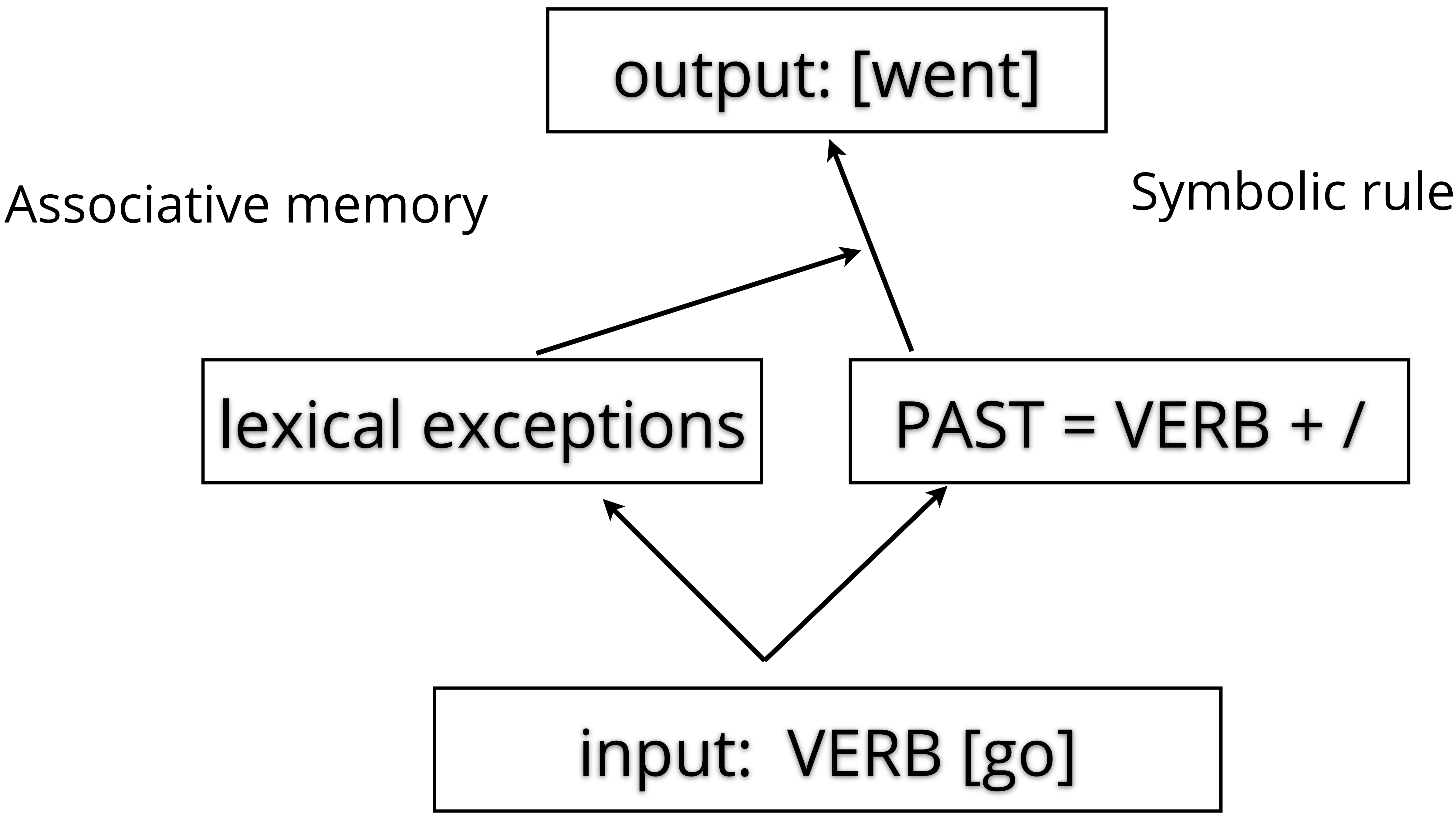
if no PAST, then PAST-TENSE:
[...X]+[ed]



loop - [lup] (PAST: [lupt])
please - [pliz] (PAST: [plizd])
grade - [gred] (PAST: [greded])
feel - [fil] (PAST: [felt])

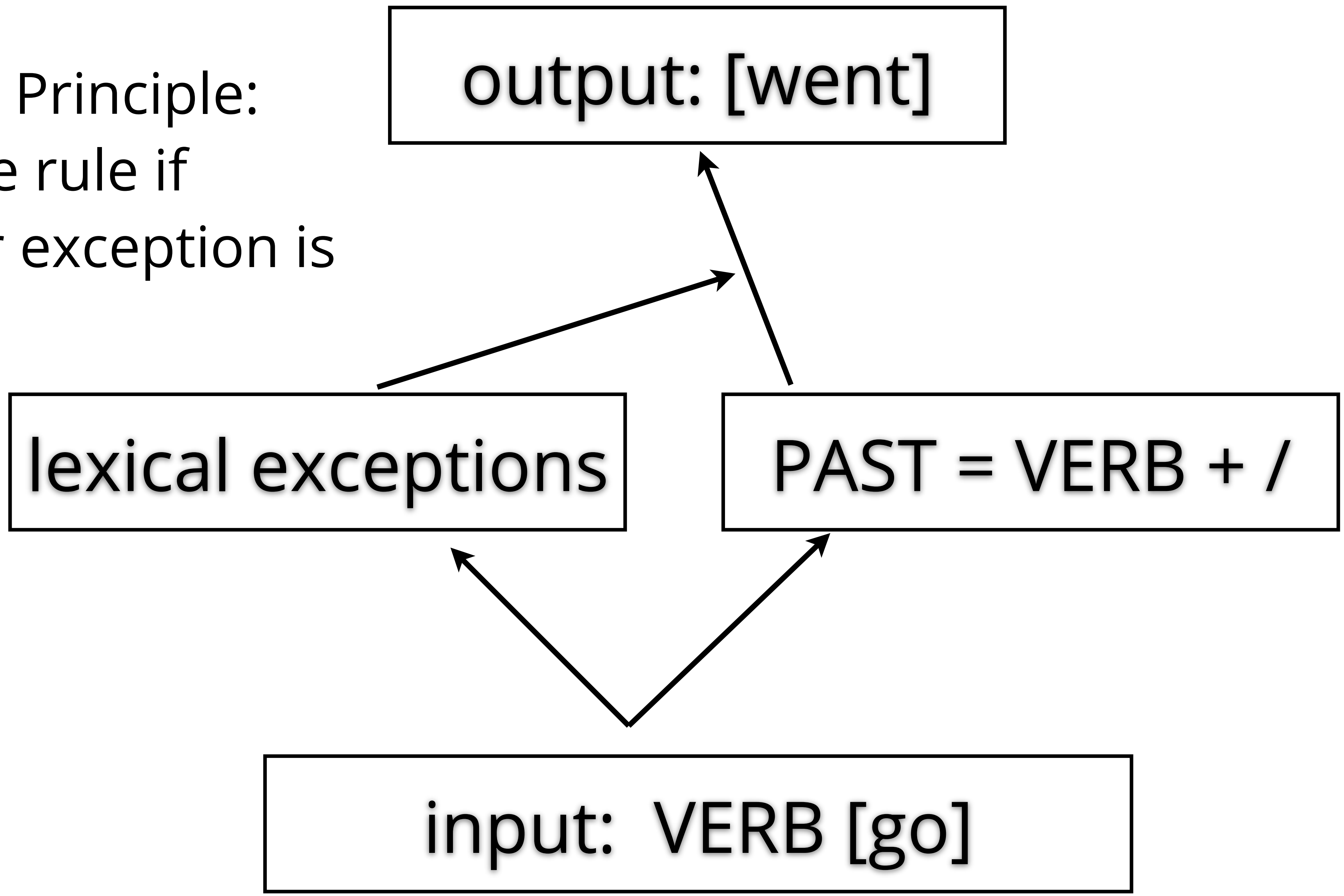


Dual route model

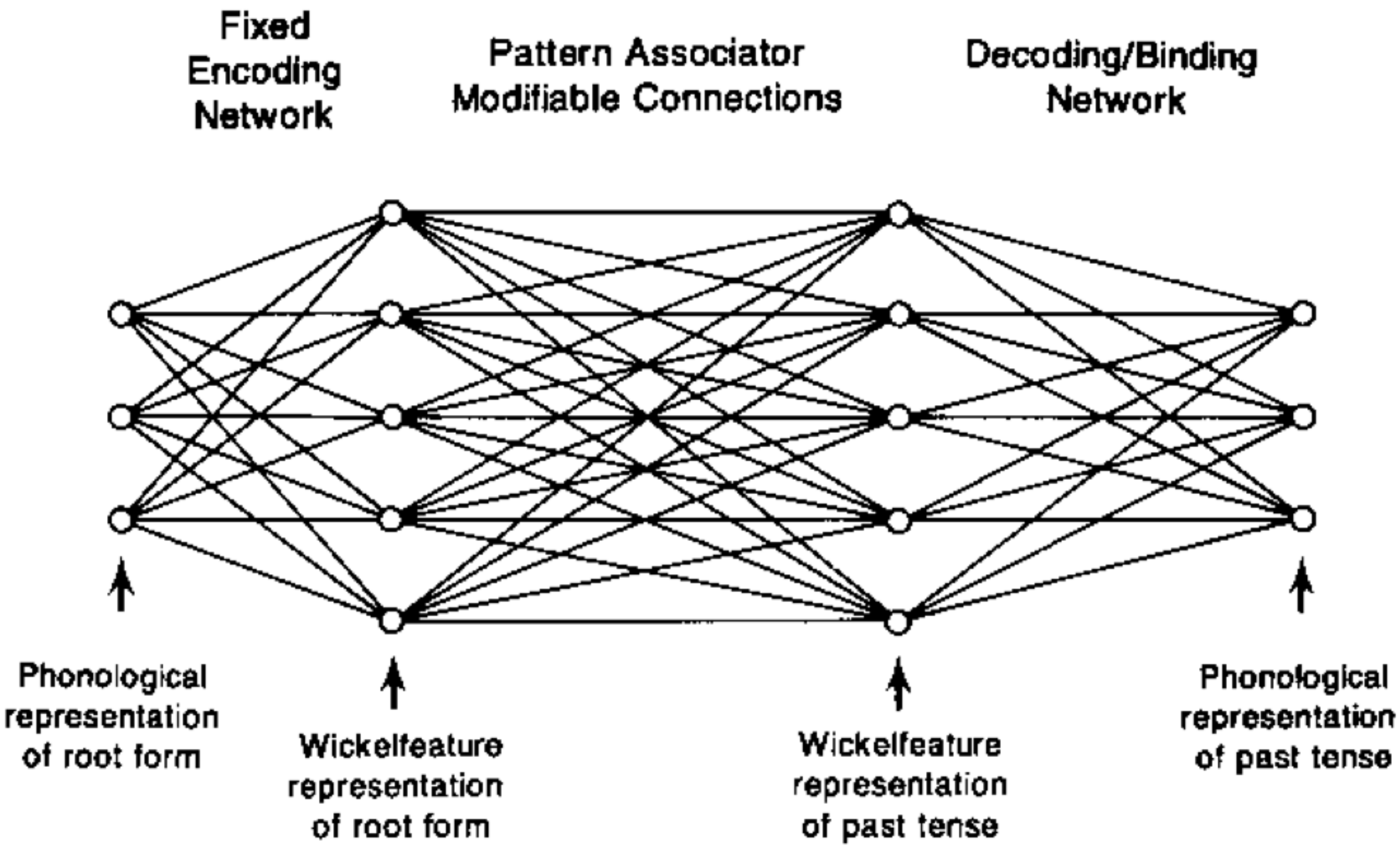


Dual route model

Blocking Principle: don’t use rule if irregular exception is listed



Connectionist model



The legacy of the PDP model

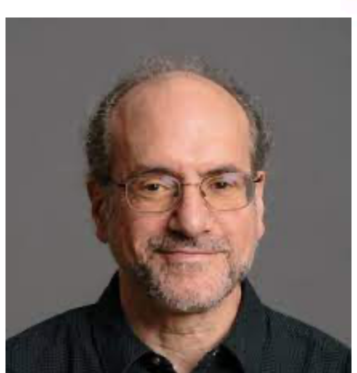
The Cognitive Science ‘Map’

East Pole cognitive theory: rooted in Rationalism

400 BC	1650	1960
Plato	Descartes Port-Royal Grammar	Carey, Spelke, Pinker Chomsky: syntax
Pāṇini		__ & Halle: phonology
” Medhatithi Gautama Mozi Aristotle	Port-Royal Logic	McCarthy: logic school of AI

West Coast theories: rooted in Empiricism and experimental science

1650 1900	1950 1975		1940 1985
Locke, Hume	Hebb	Philosophical empiricism Neuroscience	Connectionism
	Rosch	Cognitive psychology	Cognitive linguistics
Logical positivism	Skinner	Conditioning experiments	Behaviorism



From Smolensky (2020)

→ **Paul Smolensky (JHU/Microsoft)**

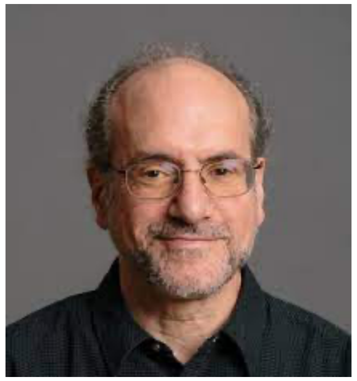
What is Cognitive Science even a science of?

→ **Knowledge (East Pole):** Characterizing the mental faculties, knowledge of individuals. Experiments, measurements of behavior, are a means to an end: Reflections of the underlying system of interest.

> **Example:** Using acceptability judgments to infer grammatical principles; Using truth value judgments in a context to infer the LFs associated with a string

→ **Behavior/Brain Dynamics (West Pole):** Characterizing how and why intelligent agents behave the way they do in a given context.

> **Example:** Neural network modeling of word by word reading times in a large dataset of reading times.



From Smolensky (2020)

The legacy of the PDP model

→ Uniform procedure: No **qualitative** distinction between regulars and irregulars.

→ Novel account of generalization.

→ Learning by exposure to examples - no hypothesis testing or explicit rule learning.

→ Domain-general. Consists of units that are not unique to language; Language in the PDP model is a reflection of more general cognitive properties.

→ No distinction between competence and performance: Knowledge arises only in the context of solving a particular task.

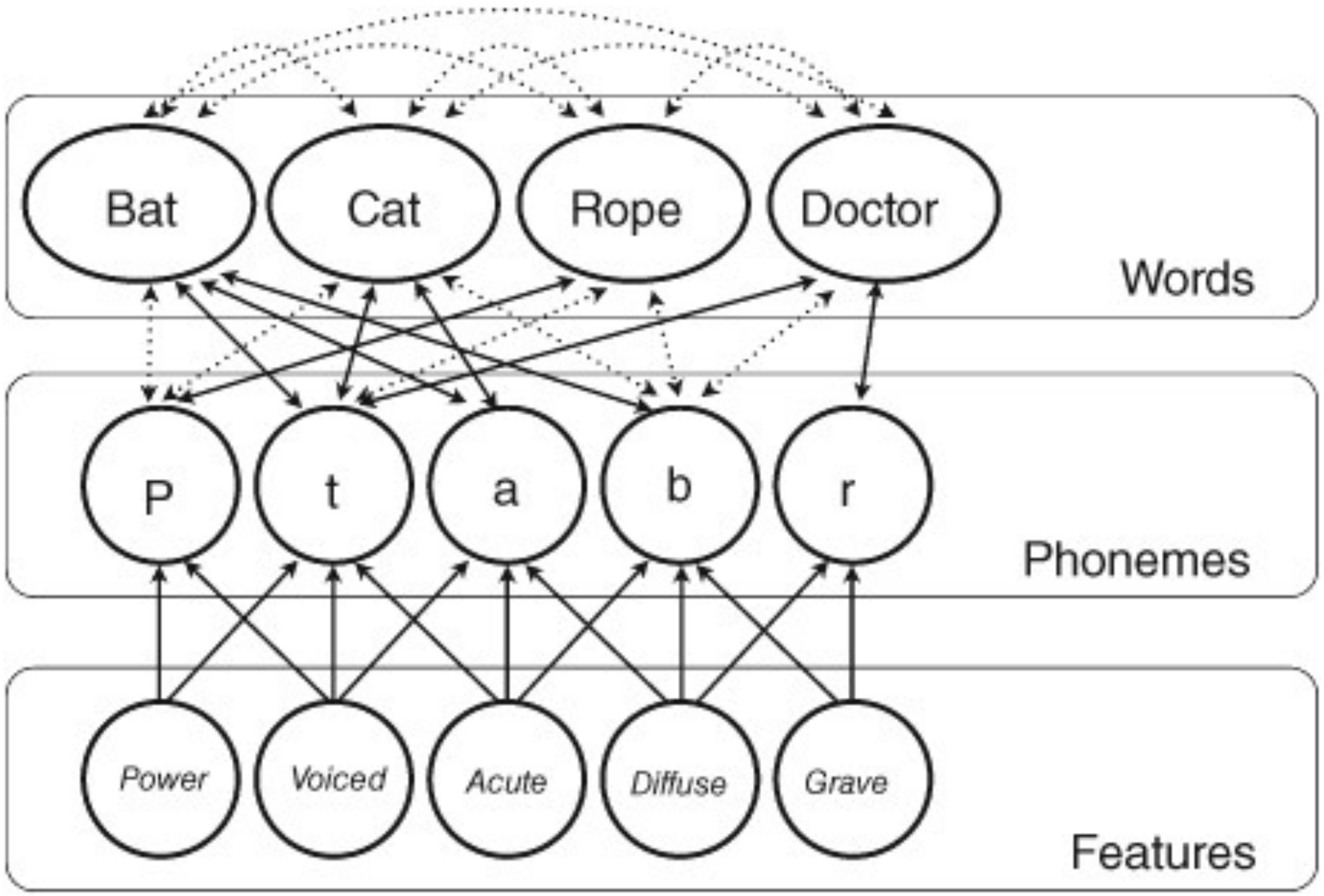
→ Understanding the nature of language requires understanding (cognitive) neurobiology, rather than analysis of primary linguistic data.

→ Computational theories are a key part of psychological theorizing.

The legacy of the PDP model

→ **Graded effects arise from the interaction of multiple soft / violable constraints!**

- Optimality Theory et al.
- TRACE model of speech recognition
- Constraint-based sentence processing (to be seen)



Issue #1: What do we need to explain?

Regulars:

- kick - kicked
- pull - pulled
- groan - groaned
- vie - vied
- introduce - introduced
- bloviate - bloviated
- ossify - ossified
- hand - handed
- yelp - yelped

Irregulars:

- sleep - slept
- ring - rang
- buy - bought
- go - went
- come - came
- eat - ate
- take - took
- bring - brought
- drink - drank

View #1: There is a ‘core’ of cases the grammar must account for, with the rest stored in a look-up table (lexicon).

Issue #1: What do we need to explain?

Regulars:	Irregulars:
kick - kicked	sleep - slept
pull - pulled	ring - rang
groan - groaned	buy - bought
vie - vied	go - went
introduce - introduced	come - came
bloviate - bloviated	eat - ate
ossify - ossified	take - took
hand - handed	bring - brought
yelp - yelped	drink - drank

View #2: There's no in principle distinction between these two types of forms - the theory must explain how speakers achieve both types.

Issue #1: What do we need to explain?

Regulars:	Irregulars:
kick - kicked	sleep - slept
pull - pulled	ring - rang
groan - groaned	buy - bought
vie - vied	go - went
introduce - introduced	come - came
bloviate - bloviated	eat - ate
ossify - ossified	take - took
hand - handed	bring - brought
yelp - yelped	drink - drank

View #2: There's no in principle distinction between these two types of forms - the theory must explain how speakers achieve both types.

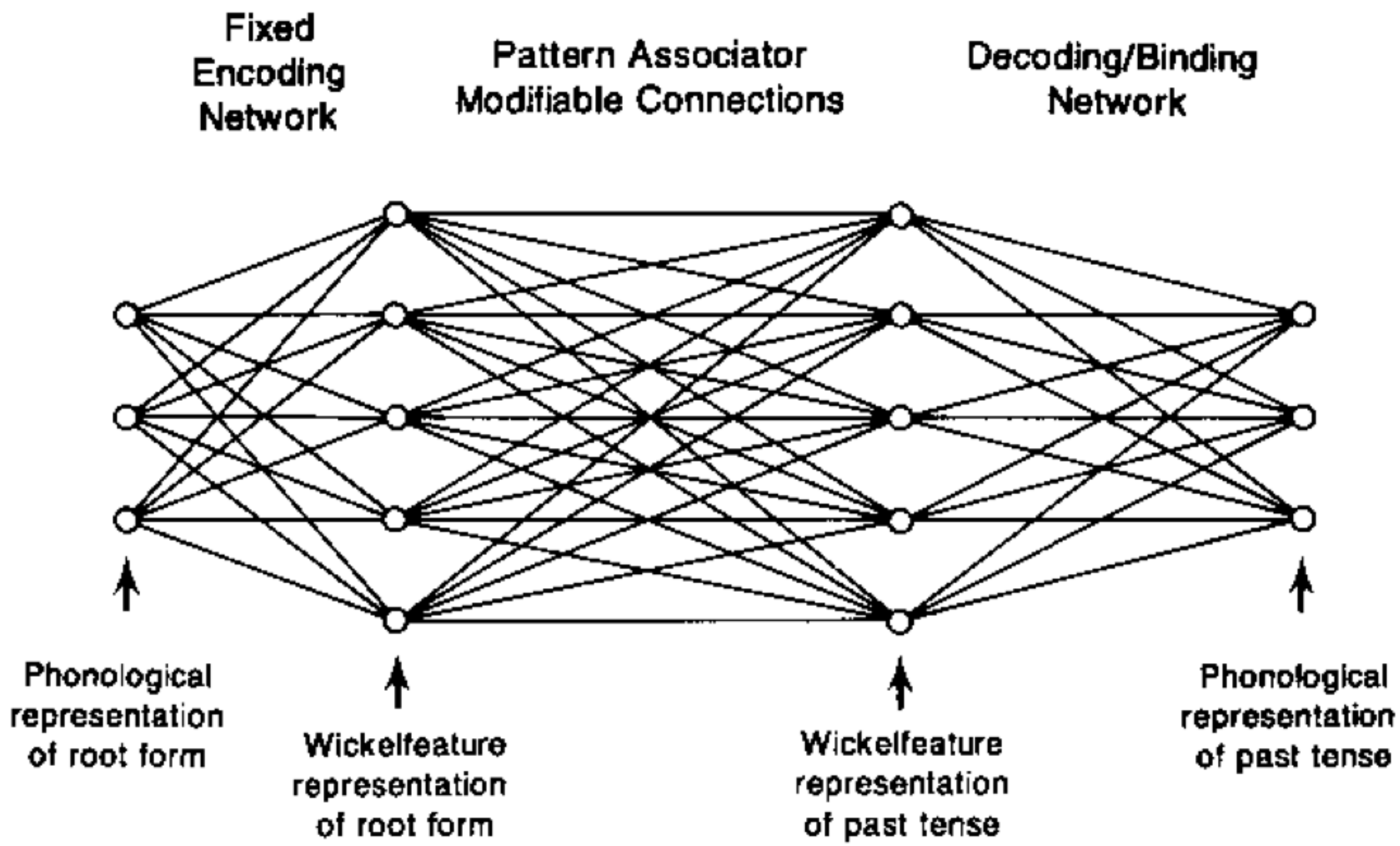
Issue #2: What makes a psycholinguistic theory?

Words and rules**:** Verbal theory. Makes commitments on central distinctions- there is an associative memory, and a rule-based mechanism. Predictions are broad/qualitative and derived by reasoning from these premises.

Rumelhart & McClelland**:** Computationally implemented theory. Makes commitments on central distinctions - there is a single pattern associator - as well as less central features, necessary to derive predictions (e.g. Wickelphones). Predictions are precise/quantitative and derived by simulation.

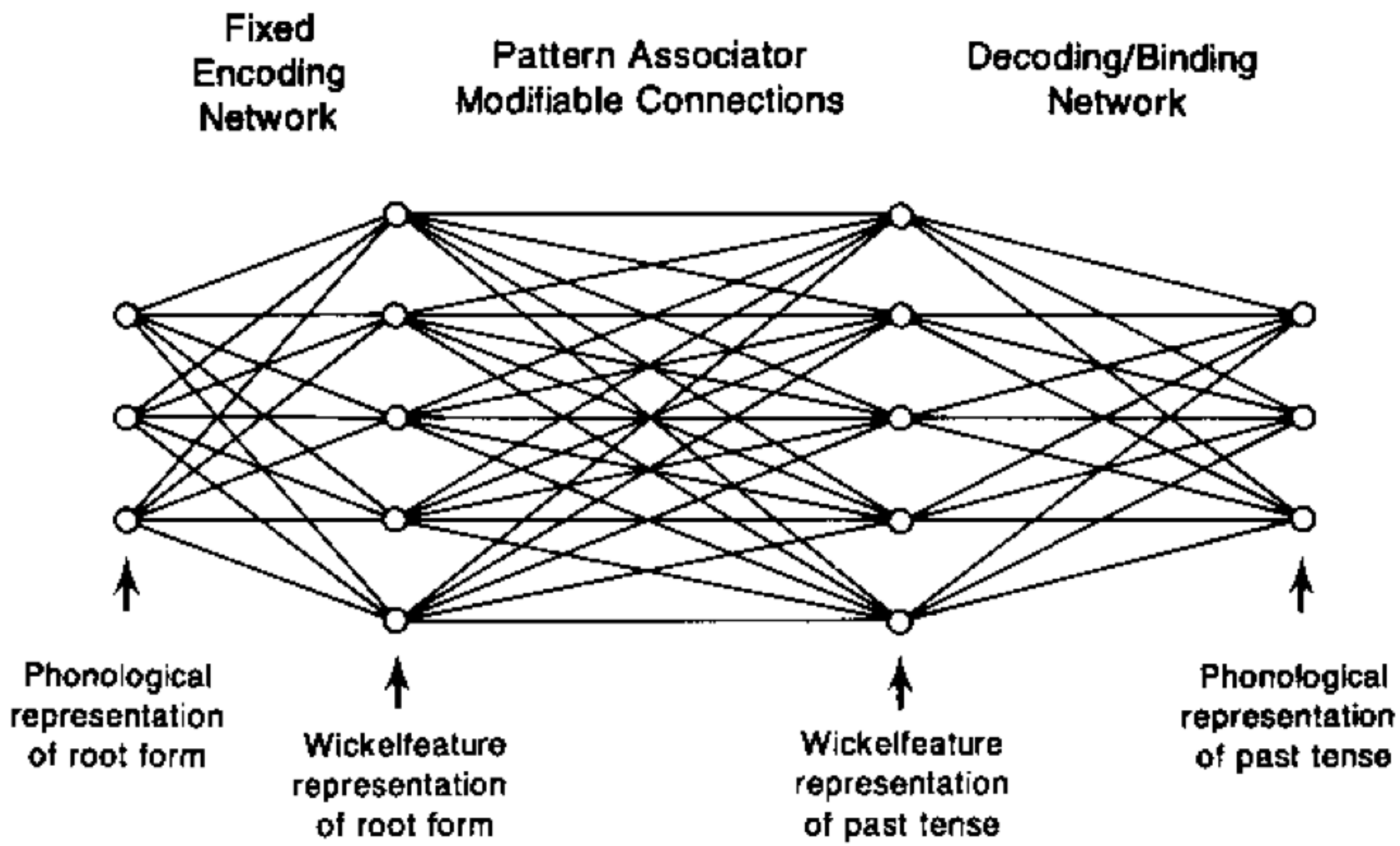
→ **What are the strengths and weaknesses of each approach? Is it a fair comparison?**

Issue #3: Competence or performance?



→ The network's knowledge lives 'in' the network of weights that achieve the input - output mapping. Competence does not have logical priority here: Understanding the task the network solves (its performance task) has priority.

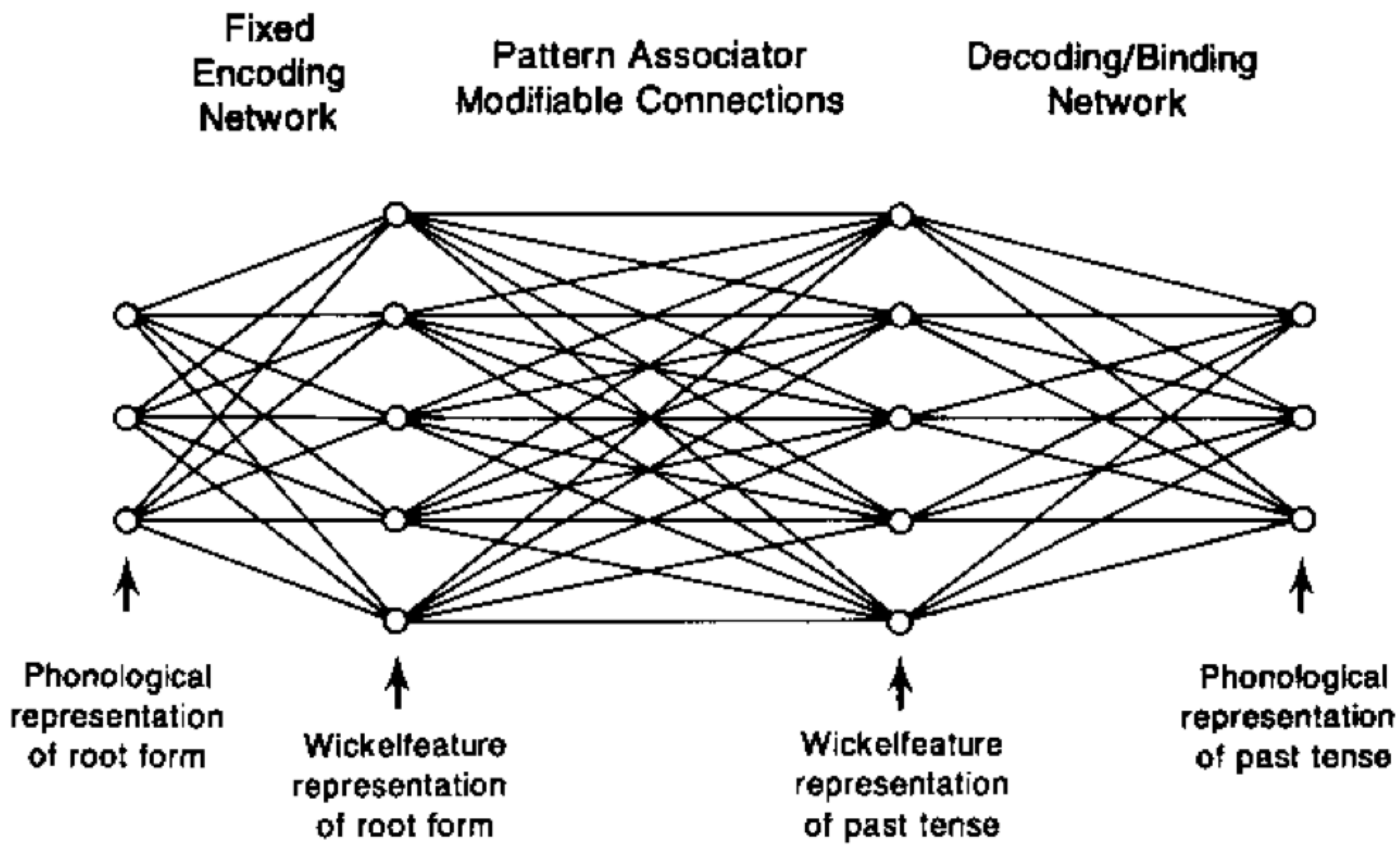
Issue #3: Competence or performance?



→ **Psycholinguistics V1:** How is grammar used in real-time? Competence takes priority, we ask how it is used in real-time.

→ **Psycholinguistics V2:** How language used and acquired? Performance takes priority, and we ask how language structure follows.

Issue #4: Learning mechanism?



→ Learning only by exposure to examples: No hypothesis testing against data.

Issue #4: Learning mechanism?

Gary Dell

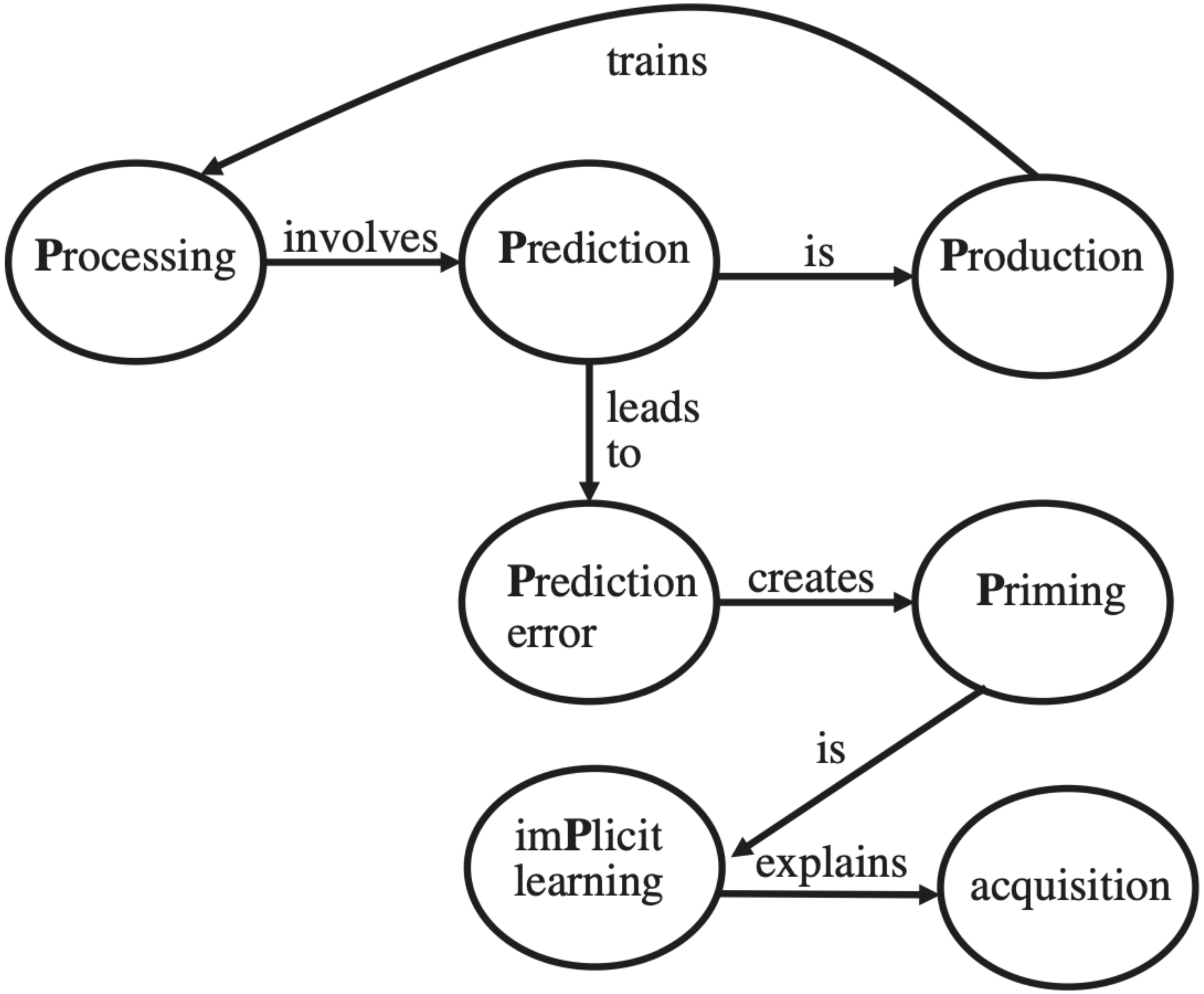


Figure 1. The P-chain framework for psycholinguistics.

Issue #4: Learning mechanism?

Charles Yang

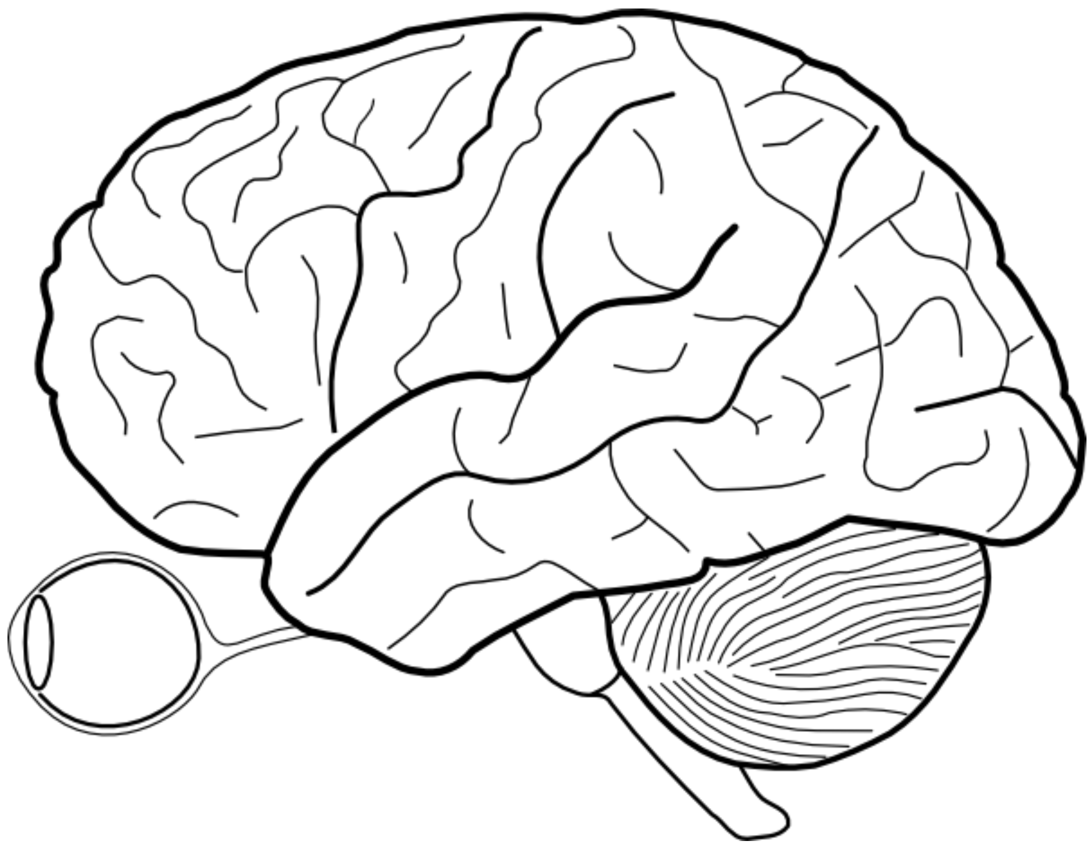
Lisa Pearl

Opinion TRENDS in Cognitive Sciences Vol.8 No.10 October 2004
Full text provided by www.sciencedirect.com
ELSEVIER SCIENCE DIRECT

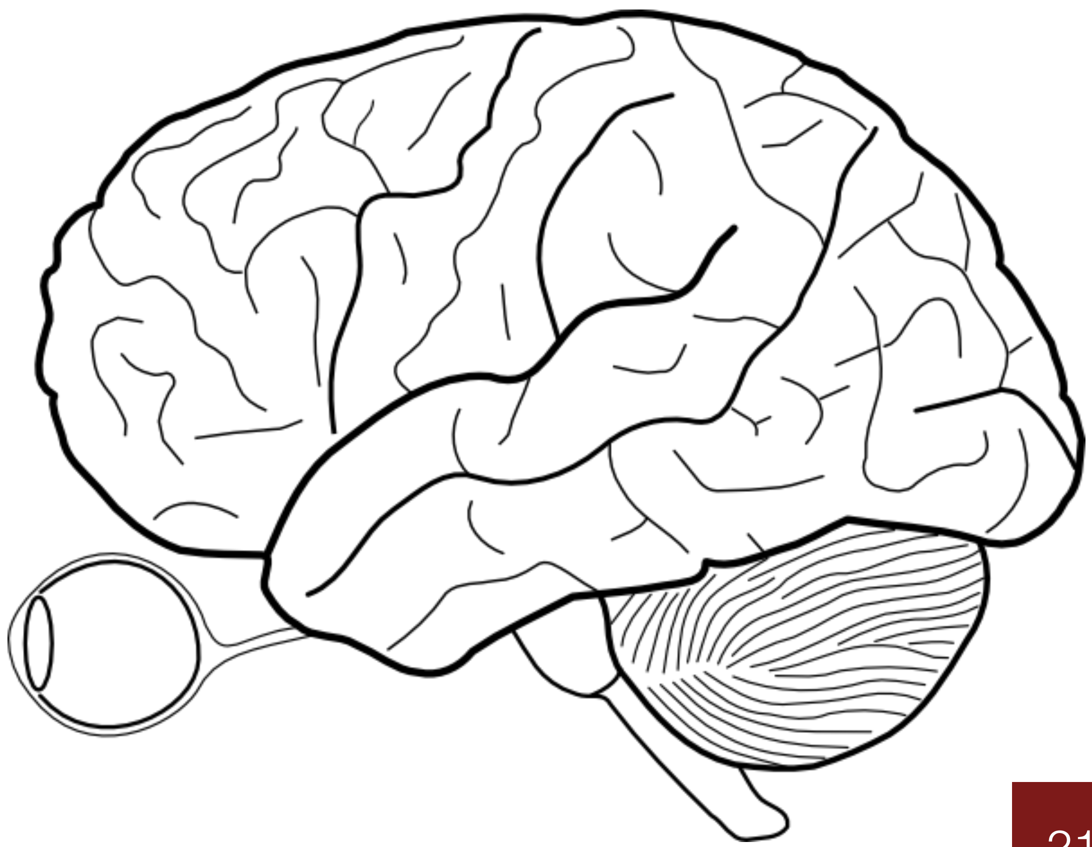
Universal Grammar, statistics or both?

Charles D. Yang

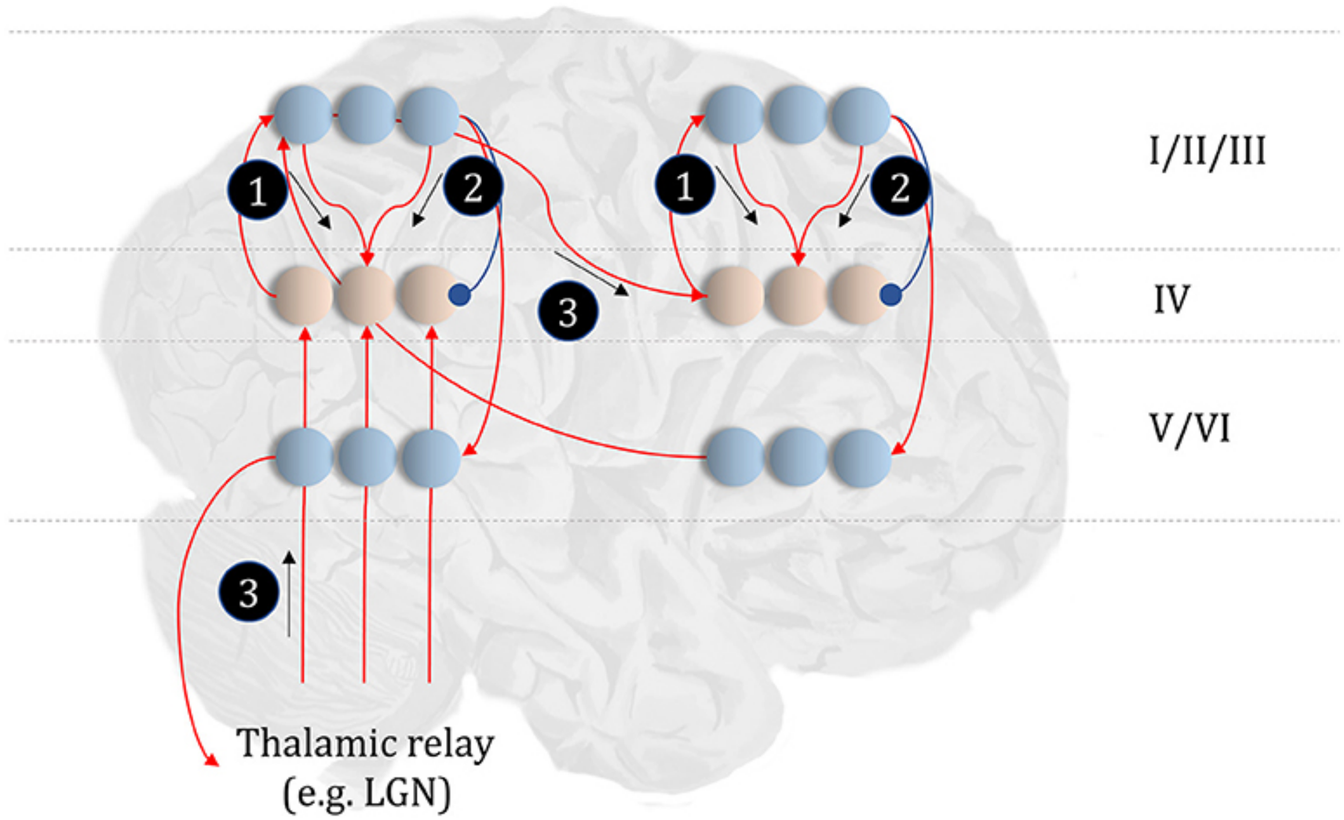
Department of Linguistics and Psychology, Yale University, 370 Temple Street 302, New Haven, CT 06511, USA

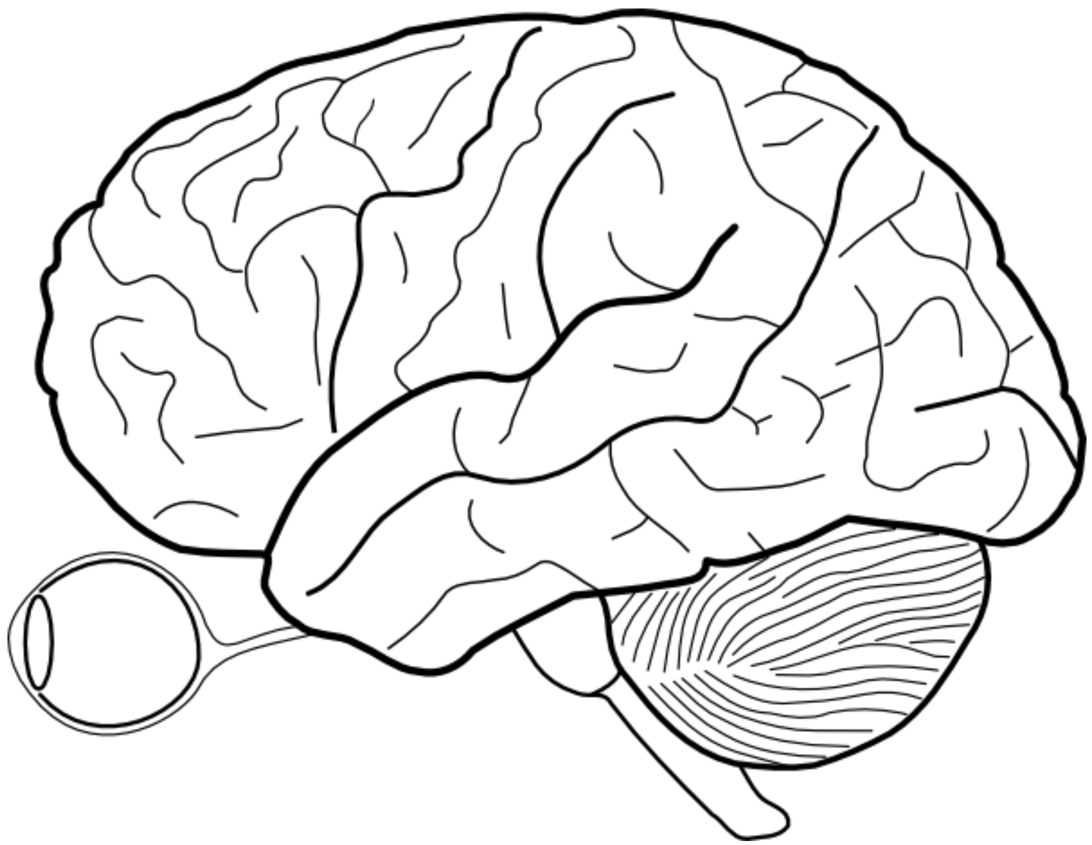


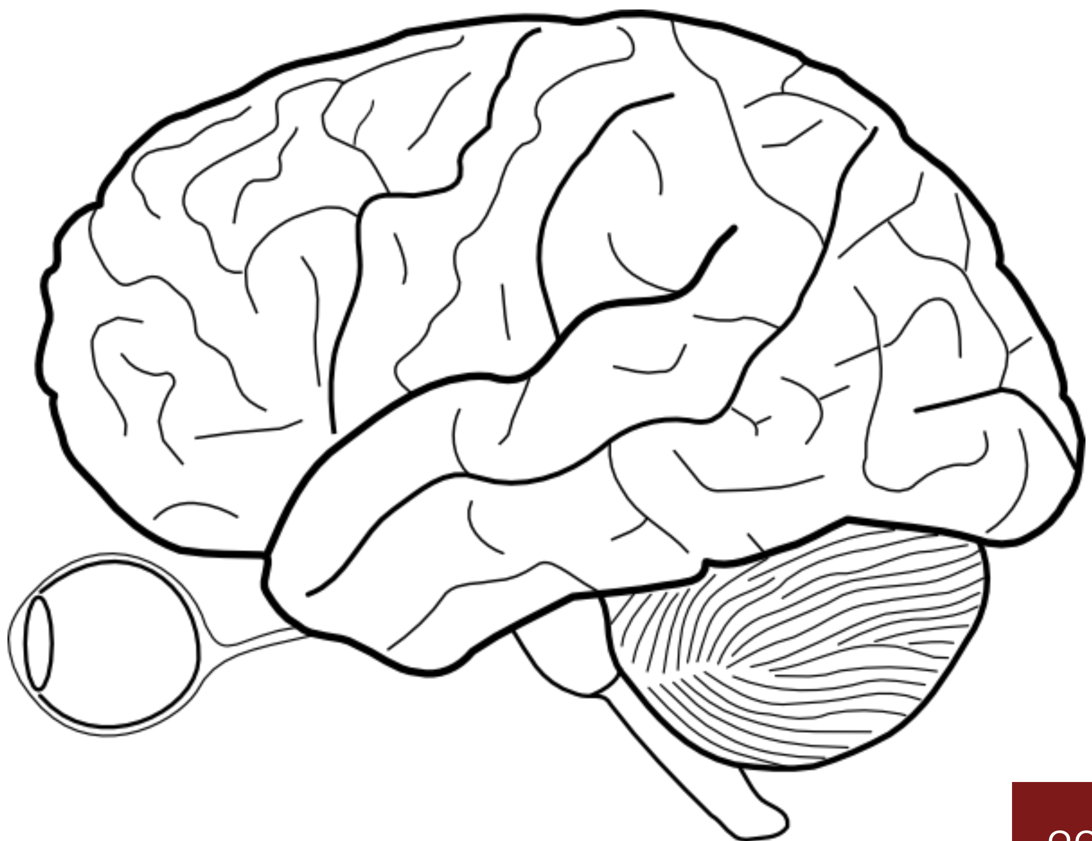
→ **If we can figure out the hardware**, then we can better understand the software that runs on it.



Issue #5: Understanding cognitive neurobiology is key to understanding language







Whatever next? Predictive brains, situated agents, and the future of cognitive science

Andy Clark
School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, EH8 9AD Scotland, United Kingdom
andy.clark@ed.ac.uk
http://www.philosophy.ed.ac.uk/people/full-academic/andy-clark.html

Abstract: Brains, it has recently been argued, are essentially prediction machines. They are bundles of cells that support perception and action by constantly attempting to match incoming sensory inputs with top-down expectations or predictions. This is achieved using a hierarchical generative model that aims to minimize prediction error within a bidirectional cascade of cortical processing. Such accounts offer a unifying model of perception and action, illuminate the functional role of attention, and may neatly capture the special contribution of cortical processing to adaptive success. This target article critically examines this "hierarchical prediction machine" approach, concluding that it offers the best clue yet to the shape of a unified science of mind and action. Sections 1 and 2 lay out the key elements and implications of the approach. Section 3 explores a variety of pitfalls and challenges, spanning the evidential, the methodological, and the more properly conceptual. The paper ends (sections 4 and 5) by asking how such approaches might impact our more general vision of mind, experience, and agency.

Keywords: action; attention; Bayesian brain; expectation; generative model; hierarchy; perception; precision; predictive coding; prediction; prediction error; top-down processing

Example: Donkey Kong on the Atari

→ Run on the MOS6502 microchip, in the Atari, Commodore 64, and Apple I.

→ Original Donkey Kong is run on this chip: the video game results directly from the actions of the chip.

→ Jonas & Kording (2017) measured the behavior of every single part of the chip as it ran Donkey Kong: 1.5 GB per second worth of data. They submitted this data to state of the art Neuroscience/Machine-learning methods for analyzing neural computation.

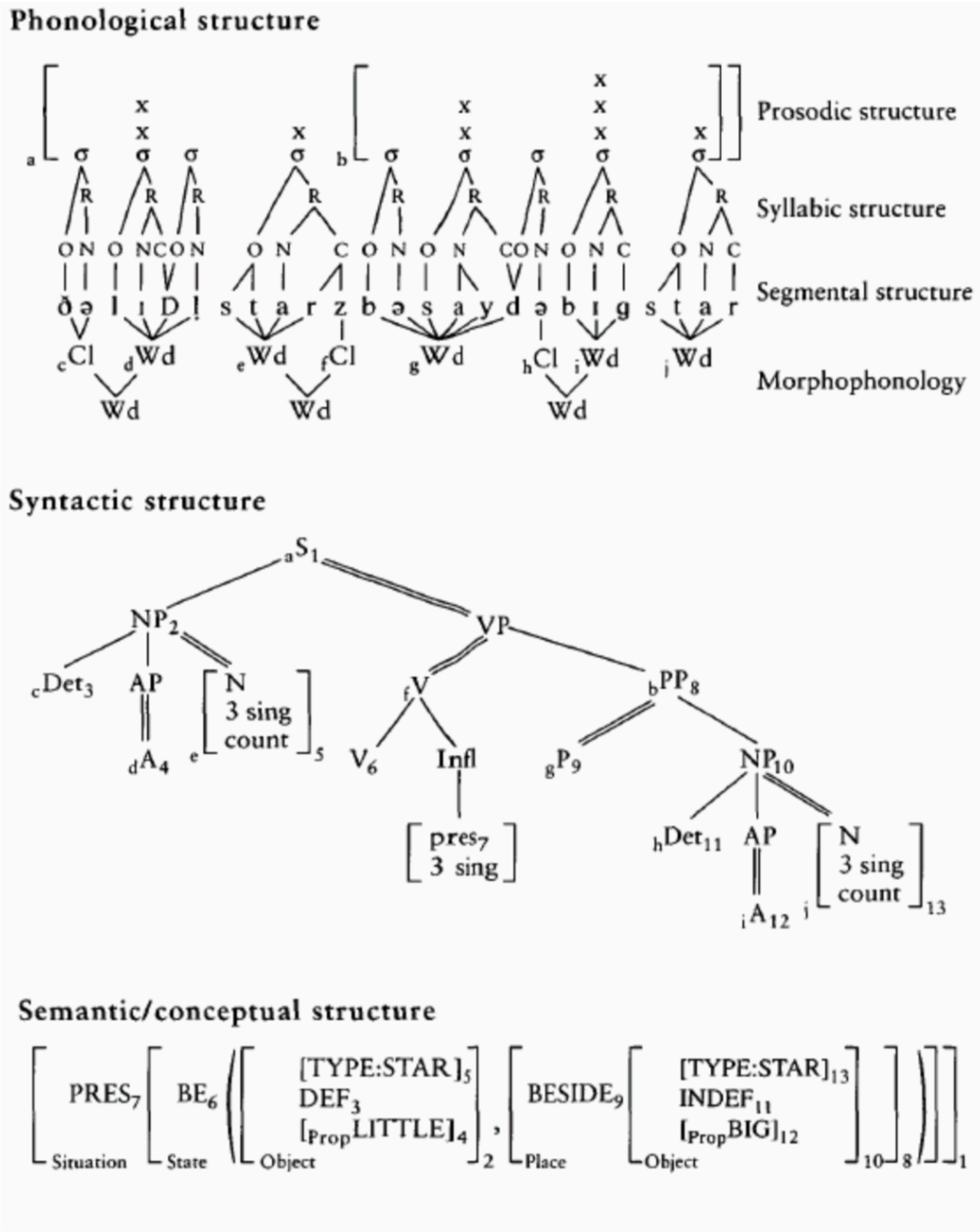
→ <u>They were unable to recover satisfactory understanding of the structure of any of the programs running</u>

Functionalism.

→ Functionalists believe that mental states can be defined functionally by the effects they can have on other mental states, or on behavior.

→ In a similar fashion, psycholinguists with a functional perspective hold that we understand the language ‘software’ by modeling it at an ‘intermediate’ level of description.

→ This stands in contrast to a **reductionist** position, which holds that it is essential to understand the brain itself in order to understand the mind, since the latter is (presumably) a result of the former.



Functionalist's response

→ We don't need to worry about this yet. We can arrive at a satisfactory understanding of the system by describing it in terms of the intermediate level of description, which contains state that functionally interact with each other (barrels, Donkey Kong, Mario, and so on).

→ So it goes with language: most psycholinguists are not focused on understanding how the brain computes. They focus on how the 'language software' runs (e.g. performance), looking to Linguistic Theory to inform them about language 'data structures' (e.g. competence).

Marr

Marr (1982): Information processing systems may be modeled/at three nested levels of abstraction:

→ **Computational level**: Description of an abstract function that maps inputs to outputs, a high-level characterization of a system's operation. **Goal is simply to understand the *what* a system does, and *why***.

→ **Algorithmic level:** Description of the *process* by which a system maps inputs to outputs, a 'medium-level' characterization of its operation. **Goal is to understand *how* (and *when*) a system accomplishes its goal.** The information, the representations it uses, and the steps by which it maps input to output.

→ **Implementational level:** Description of the machine/circuit that implements an algorithm, a 'low-level' characterization of a systems operation.

Marr

Example: **Cash Register**.

Computational Level (What?): The machine is used to sum the items purchased. More specifically, it maps a list of values for a list individual items to a total cost of that list. The specific function is addition: total due = cost(item 1)+cost(item 2)+ …

Algorithmic Level (By what method?): Representations/data structures need to be specified (Arabic numerals? Roman numerals? Binary?), and processes need to be specified to transit from input to output. Example: If you use Arabic numerals, you can add least significant digits first, carry a 1 if result > 9, move left one digit, and repeat until terminating. The algorithm is slightly different for binary numerals (same but carry a 1 if result > 2). Good luck if you're using roman numerals!

Implementation Level: (How is this carried out physically?): Circuits for electronic registers, or brass wheels that rotated kept track of values in the dollars and cents positions for older models.

Another Drosophila?

- **Do neural networks learn syntax-sensitive dependencies?** English subject-verb agreement as a test case (Linzen, Dupoux & Goldberg, 2016):

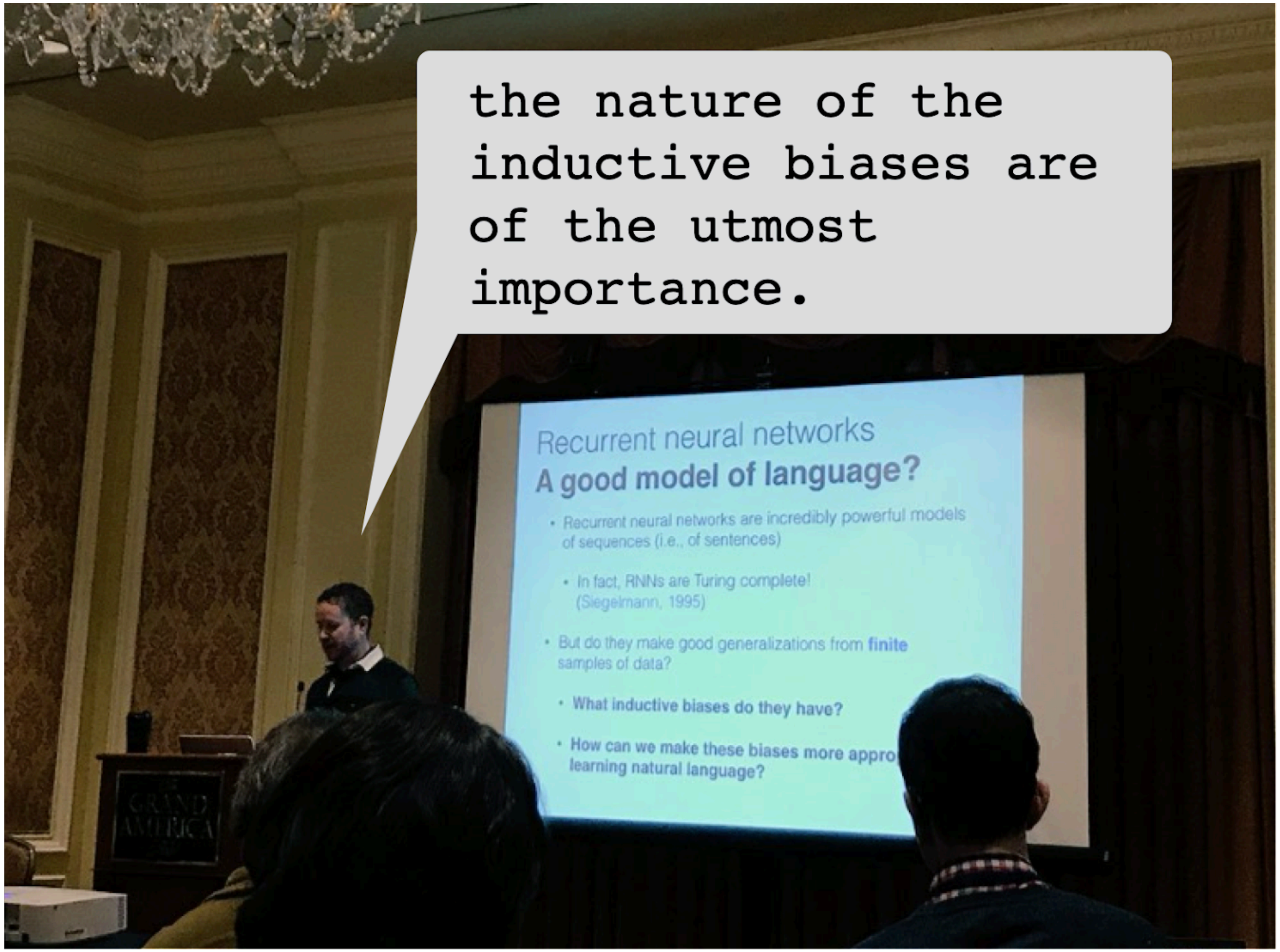
the key is / *are on the table

the key to the cabinets is / *are on the table

the children said **the key** to the cabinets is / *are on the table

the key to the cabinets that I almost destroyed is / *are on the table

Another Drosophila?



Agreement attraction: Chris Dyer @ SCiL 2018

Another Drosophila?

`Which flowers are the gardeners planting?`

Tanner et al (2014)

Table 4
Example experimental sentences from Experiment 3.

Attractor number	Verb type	Sentence
Singular	Agreeing	*The chemist with the test tube is conducting an experiment*
Singular	Modal	*The chemist with the test tube might be conducting an experiment*
Plural	Agreeing	*The chemist with the test tubes is conducting an experiment*
Plural	Modal	*The chemist with the test tubes might be conducting an experiment*

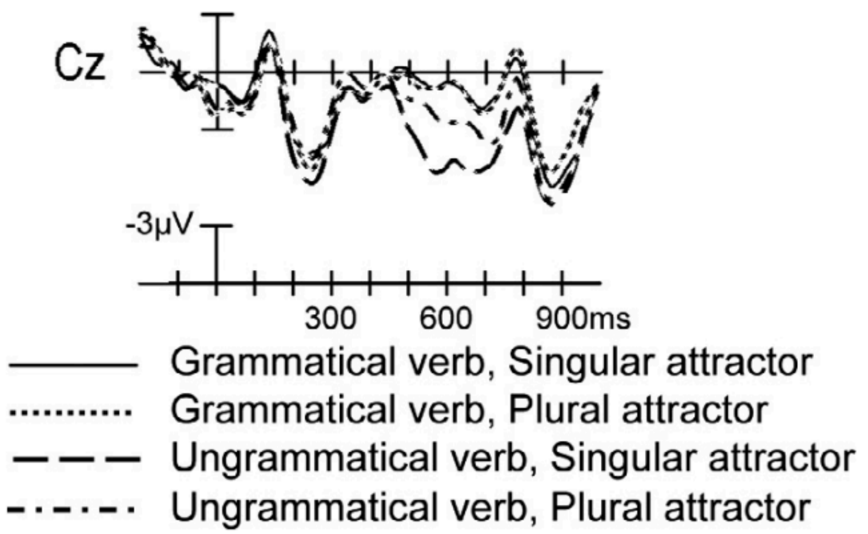


Fig. 3. Grand mean ERP waveforms for all four experimental conditions in Experiment 1 at midline vertex electrode Cz. Onset of the verb is indicated by the vertical calibration bar; each tick mark represents 100 ms of time. Negative voltage is plotted up.

- **P600:** Posterior postivity that peaks around 600ms (broadly) post-stimulus. Triggered by ungrammatical words, difficult to integrate words (e.g. words in a long-distance syntactic dependency), garden path sentences, and more…

Tanner et al (2014)

Table 4
Example experimental sentences from Experiment 3.

Attractor number	Verb type	Sentence
Singular	Agreeing	*The chemist with the test tube is conducting an experiment*
Singular	Modal	*The chemist with the test tube might be conducting an experiment*
Plural	Agreeing	*The chemist with the test tubes is conducting an experiment*
Plural	Modal	*The chemist with the test tubes might be conducting an experiment*

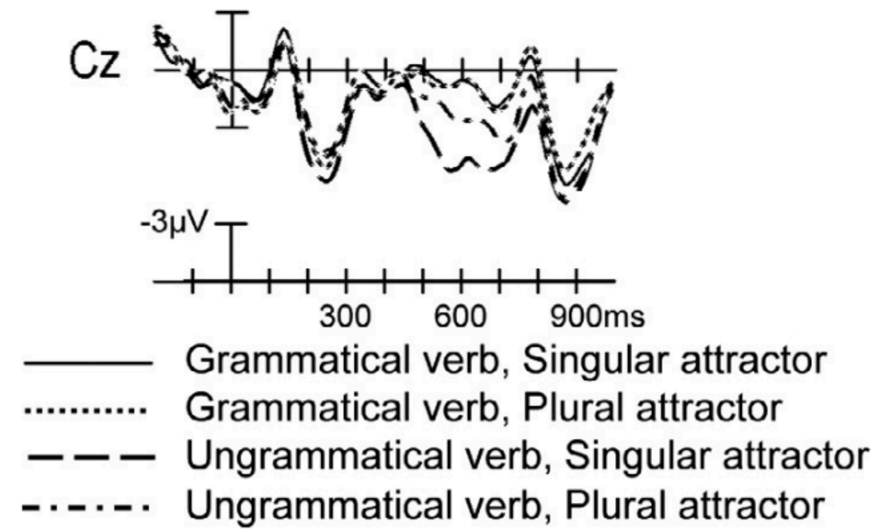


Fig. 3. Grand mean ERP waveforms for all four experimental conditions in Experiment 1 at midline vertex electrode Cz. Onset of the verb is indicated by the vertical calibration bar; each tick mark represents 100 ms of time. Negative voltage is plotted up.

View #1: There is a 'core' of cases the grammar must account for (normal, structured agreement), with the rest relegated to performance mechanisms (errors caused by memory, perhaps)

Tanner et al (2014)

Table 4
Example experimental sentences from Experiment 3.

Attractor number	Verb type	Sentence
Singular	Agreeing	*The chemist with the test tube is conducting an experiment*
Singular	Modal	*The chemist with the test tube might be conducting an experiment*
Plural	Agreeing	*The chemist with the test tubes is conducting an experiment*
Plural	Modal	*The chemist with the test tubes might be conducting an experiment*

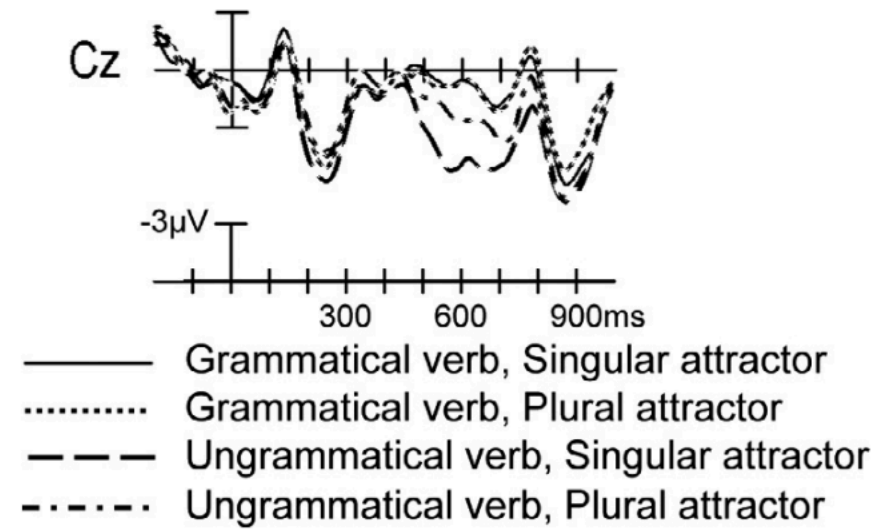


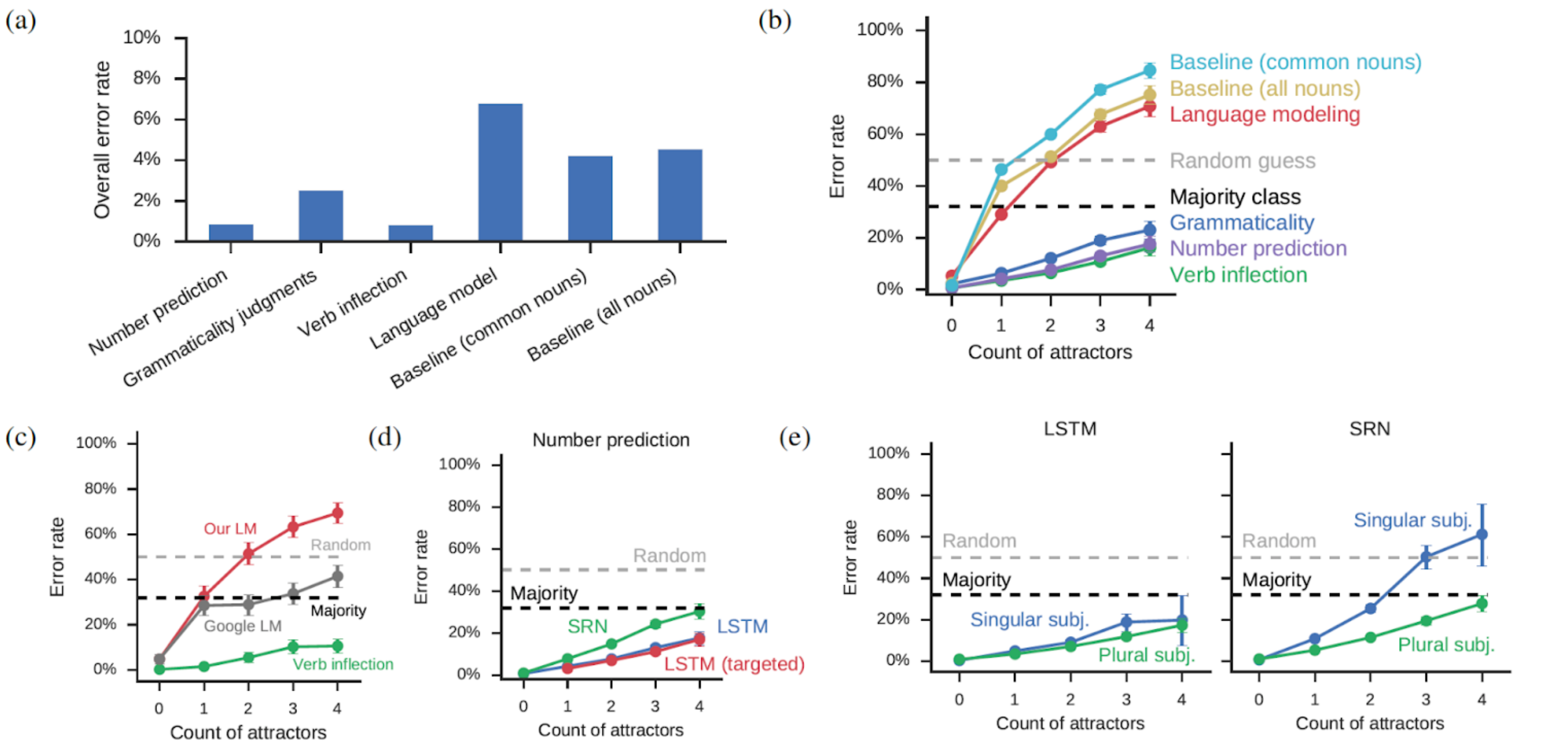
Fig. 3. Grand mean ERP waveforms for all four experimental conditions in Experiment 1 at midline vertex electrode Cz. Onset of the verb is indicated by the vertical calibration bar; each tick mark represents 100 ms of time. Negative voltage is plotted up.

View #2: There's no in principle distinction between these two types of agreement behavior - the theory must explain how speakers arrive at both types

Another Drosophila?

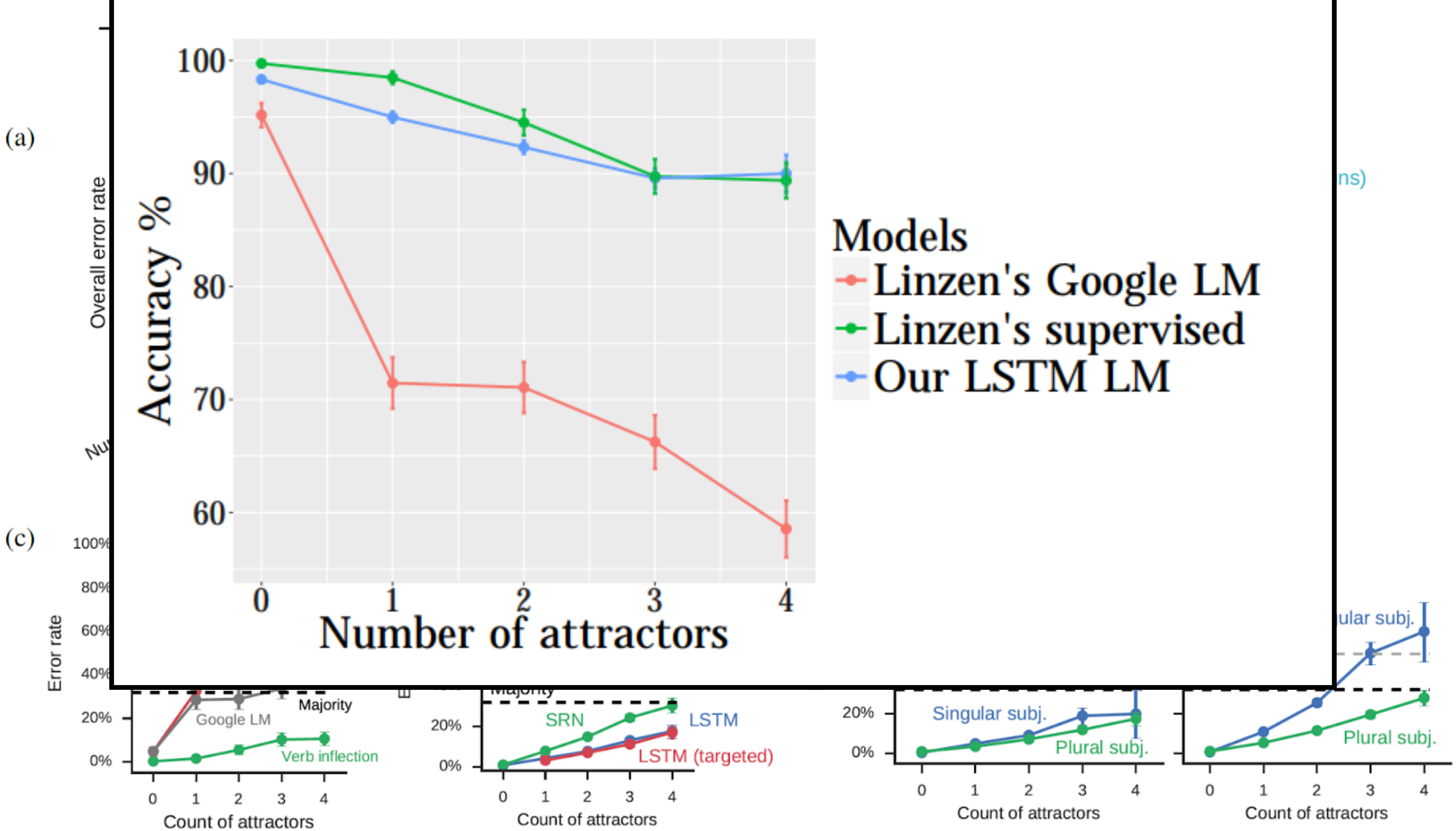
Training objective	Sample input	Training signal	Prediction task	Correct answer
Number prediction	*The keys to the cabinet*	PLURAL	SINGULAR/PLURAL?	PLURAL
Verb inflection	*The keys to the cabinet [is/are]*	PLURAL	SINGULAR/PLURAL?	PLURAL
Grammaticality	*The keys to the cabinet are here.*	GRAMMATICAL	GRAMMATICAL/UNGRAMMATICAL?	GRAMMATICAL
Language model	*The keys to the cabinet*	are	$P(are) > P(is)$?	True

Table 1: Examples of the four training objectives and corresponding prediction tasks.



Another Drosophila?

Training objective	Sample input	Training signal	Prediction task	Correct answer
Number prediction	*The keys to the cabinet*	PLURAL	SINGULAR/PLURAL?	PLURAL
Verb inflection	*The keys to the cabinet [is/are]*	PLURAL	SINGULAR/PLURAL?	PLURAL
Grammaticality	*The keys to the cabinet are here*	GRAMMATICAL	GRAMMATICAL/UNGRAMMATICAL?	GRAMMATICAL



Pinker & Prince (1988)

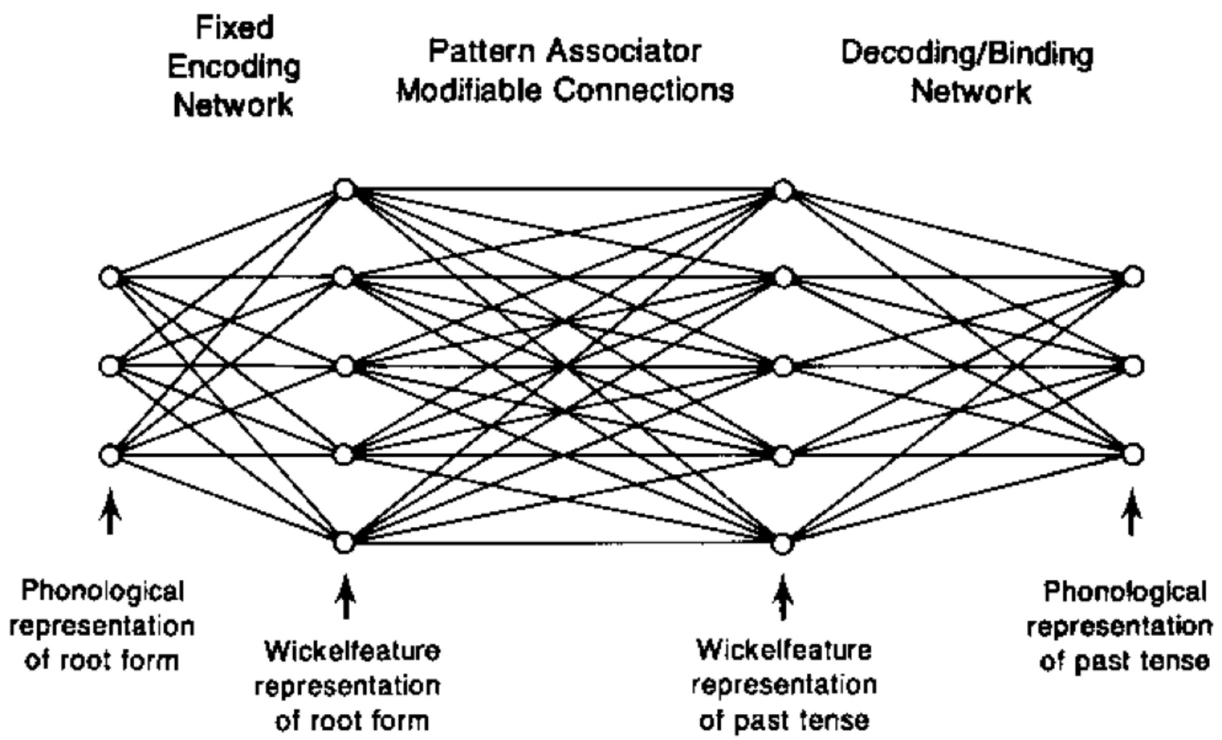
- Unable to account for linguistic structure (denominal verbs are regular: *ringed* vs *rang*)

- Unable to account for differences in compounding: *mice-eater* but not **rats-eater*.

- Unable to account for homophony *break* vs *brake*

- Unable to represent temporal order (*rapata* vs *ratapa*)

- Predicts that there should be typicality effects for regulars and irregulars alike.



Pinker & Prince (1988)

→ **Predictions of an abstract rule:**

- There should be no effect of stem typicality: The abstract rule should apply equally well no matter what value supplies the variable.
- The rule is a regular **elsewhere** case: It applies when there is not a form that lexical memory supplies. It does not need to be the most frequent exponent / morpheme.

German plurals

	Masculine	Feminine	Neuter
Common	[-e] [¨-e] [-]	[-en] [-n] [-nen]	[-e] [-]
Less common	[-en] [-n] [¨]	[¨-e]	[-er] [¨ -er]
Adopted foreign words	[-s]		

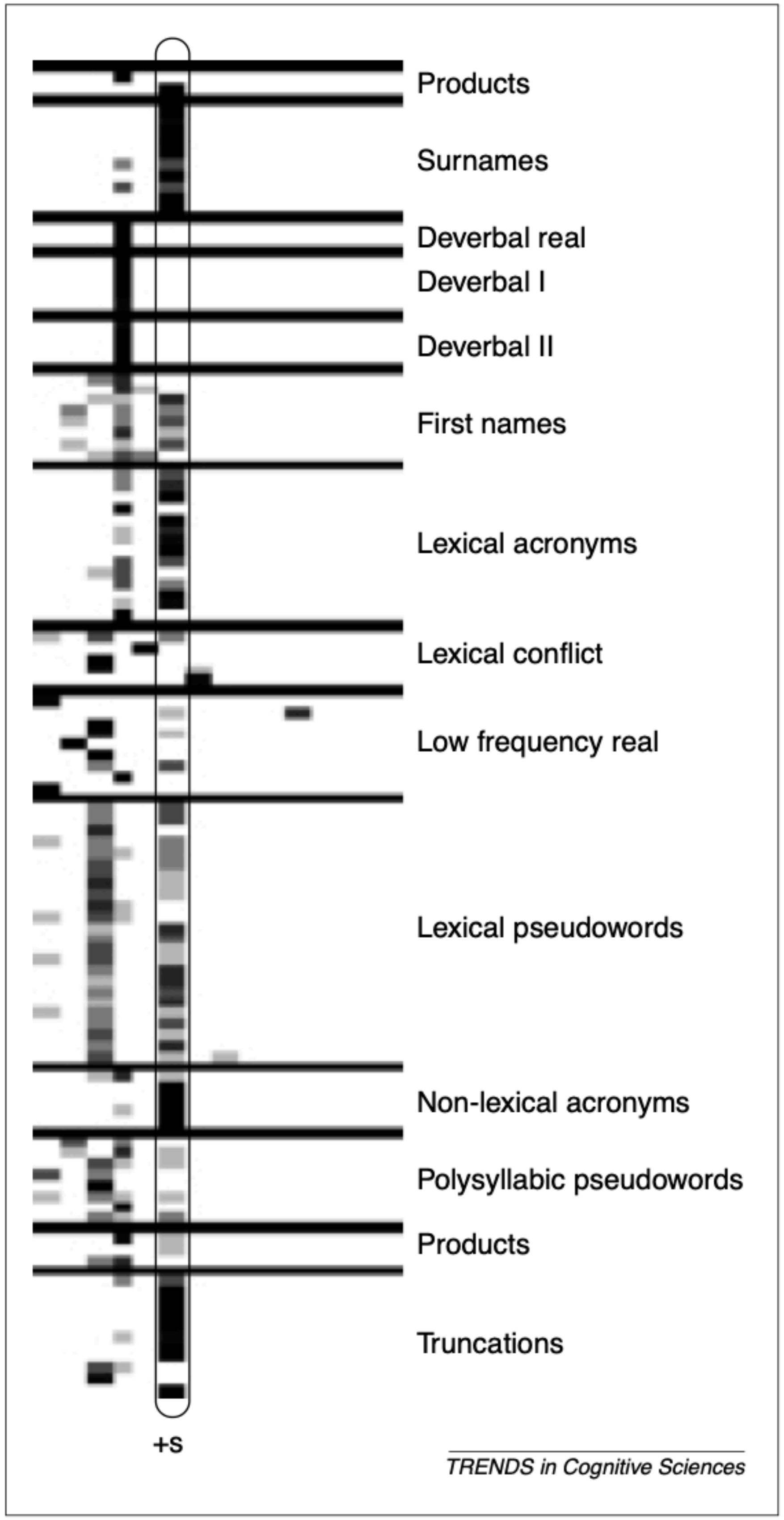
singular	plural
das Haus (the house)	die Häuser (the houses)
der Student (the student)	die Studenten (the students)
die Hand (the hand)	die Hände (the hands)
das Hobby (the hobby)	die Hobbys (the hobbies)
die Mutti (the momma)	die Muttis (the mommies)

- Applies to ~7% of nouns.
- But is (over)extended in childhood, and productively applied to unusual nouns, and 'exocentric' nouns (as in e.g. English denominal ring)

German plurals

	Masculine	Feminine	Neuter
Common	[-e] [¨-e] [-]	[-en] [-n] [-nen]	[-e] [-]
Less common	[-en] [-n] [¨]	[¨-e]	[-er] [¨ -er]
Adopted foreign words		[-s]	

singular	plural
das Haus (the house)	die Häuser (the houses)
der Student (the student)	die Studenten (the students)
die Hand (the hand)	die Hände (the hands)
das Hobby (the hobby)	die Hobbys (the hobbies)
die Mutti (the momma)	die Muttis (the mommies)



Generalization

Usually I [zek], but yesterday I [zekt]

Present Tense	***-ed***	**Vowel Change**	**No Change**
1. [zek]	a. [zekt]	b. [zʊk]	c. [zek]
2. [zet]	a. [zetəd]	b. [zʊt]	c. [zet]
3. [zep]	a. [zept]	b. [zʊp]	c. [zep]
4. [zik]	a. [zikt]	b. [zʊk]	c. [zik]
5. [zit]	a. [zitəd]	b. [zʊt]	c. [zit]

Albright and Hayes (2003)

→ **\`Islands of reliability\`:**

> “`We will refer to phonological contexts in which a particular morphological change works especially well in the existing lexicon as "Islands of Reliability”"` (p. 127)

→ Their search of the English lexicon reveals islands of reliability for both regulars and irregulars:

- **Irregulars:** ɪŋ (e.g. fling, spring, ring) is highly reliable for

$$\text{ɪ} \rightarrow \text{ʌ} \,/\, \_\_\_\, \text{ŋ}]_{[+\text{past}]}$$

- **Regularls:** all 352 stems ending in a voiceless fricative are regular.

Albright and Hayes (2003)

Table 3
Design of the Core set of wug stems

Stem occupies an island of reliability for both the regular output and at least one irregular output.	Stem occupies an island of reliability for the regular output only.
Stem occupies an island of reliability for at least one irregular output, but not for the regular output.	Stem occupies no island of reliability for either regular or irregular forms.

(14) a. Island of reliability for both regulars and irregulars
dize [daɪz] (*doze* [doz]); ***fro*** [fro] (*frew* [fru]); ***rife*** [raɪf] (*rofe* [rof], *riff* [rɪf])

b. Island of reliability for regulars only[12]
bredge [brɛdʒ] (*broge* [brodʒ]); ***gezz*** [gɛz] (*gozz* [gaz]); ***nace*** [nes] (*noce* [nos])

c. Island of reliability for irregulars only
fleep [flip] (*flept* [flɛpt]); ***gleed*** [glid] (*gled* [glɛd], *gleed*); ***spling*** [splɪŋ] (*splung* [splʌŋ], *splang* [splæŋ])

d. Island of reliability for neither regulars nor irregulars
gude [gud] (*gude*); ***nung*** [nʌŋ] (*nang* [næŋ]); ***preak*** [prik] (*preck* [prɛk], *proke* [prok])

Albright and Hayes (2003)

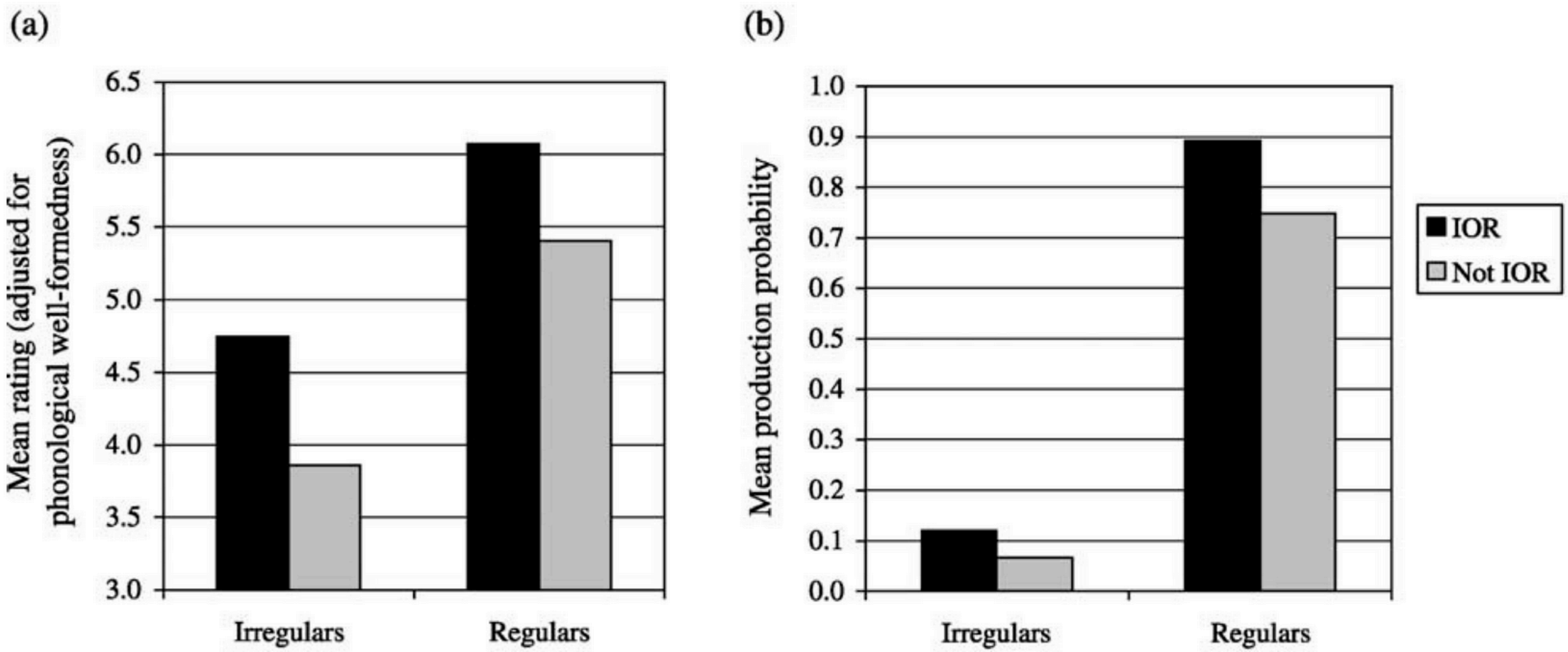


Fig. 2. Effect of islands of reliability (IOR) for irregulars and regulars. (a) IOR effect on ratings (adjusted). (b) IOR effect on production probabilities.

Generalization: Rules vs Examples

→ **Phonological rule:** Highlights **structured** similarity, similarity based on a particular structural description / rule format.

$$\text{ɪ} \longrightarrow \text{ʌ} \;/\; \underline{\quad\quad}\; \text{ŋ}]_{[+\text{past}]}$$

→ **Analogy:** Highlights **variegated** similarity, similarity based on any arbitrary aspect of similarity to another token.

Model form	s	p	l	ɪ	ŋ
fling-flung		f	l	ɪ	ŋ
sting-stung	s	t		ɪ	ŋ
"plip"-"plup"		p	l	ɪ	p
"sliff"-"sluff"	s		l	ɪ	f

Exemplar theory

→ **Generalized Context Model (Nosofsky, 1986):** An example of an **exemplar theory** of categorization: Categories are represented through examples in memory, rather than abstract summary statistics, prototypes, or other compact descriptions of the category structure.

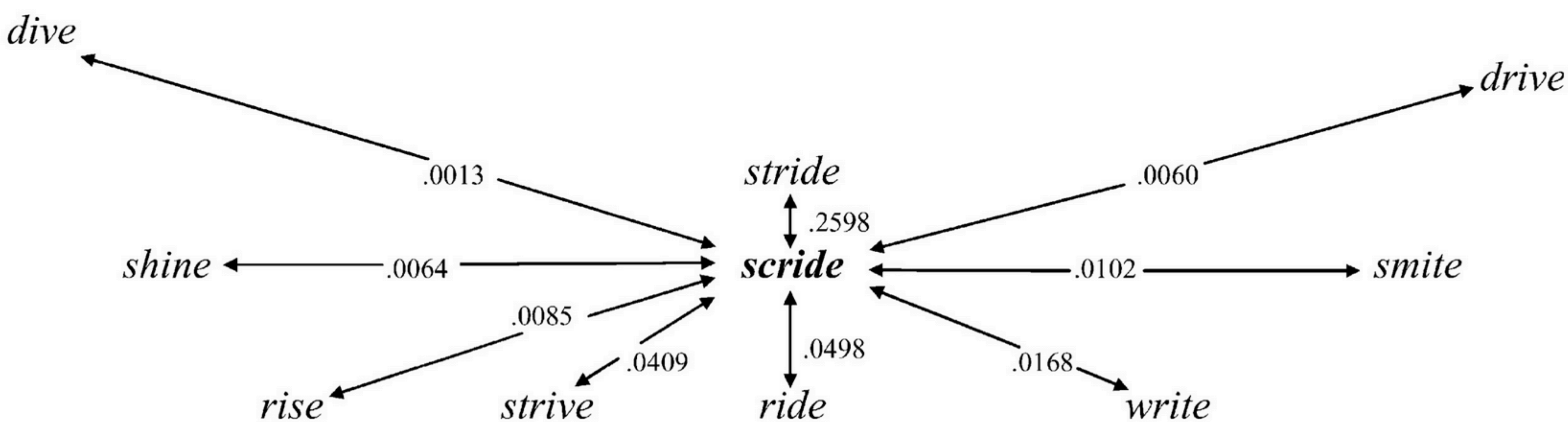


Fig. 1. Similarity of all [aɪ] → [o] forms to *scride*.

Rules all the way down

→ **Minimal Generalization Learner (Albright & Hayes, 2003):** Learns the most specific phonological rule necessary to capture a given alteration, and its reliability/validity. The grammar consists of rules all the way down - from rules that govern a single verb to extremely general (regular) rules:

a.	Change	Variable	Shared features	Shared segments	Change location		
b.	$\varnothing \rightarrow$ d / [		ʃ	aɪn	___	$]_{[+\text{past}]}$	(*shine-shined*)
c.	$\varnothing \rightarrow$ d / [	kən	s	aɪn	___	$]_{[+\text{past}]}$	(*consign-consigned*)
d.	$\varnothing \rightarrow$ d / [	X	$\begin{bmatrix}+\text{strident}\\+\text{continuant}\\-\text{voice}\end{bmatrix}$	aɪn	___	$]_{[+\text{past}]}$	(generalized rule)