Interactivity and competition: Spoken word recognition

Brian Dillon
Shota Momma

University of Massachusetts, Amherst
Department of Linguistics

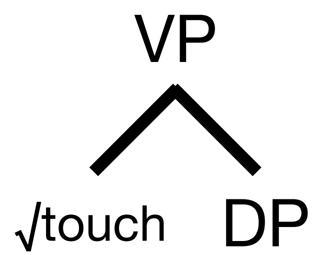
2/24/2021

Words

Today's question: How do listeners coordinate multiple levels of representation during real-time linguistic analysis?

Test case: Spoken word recognition

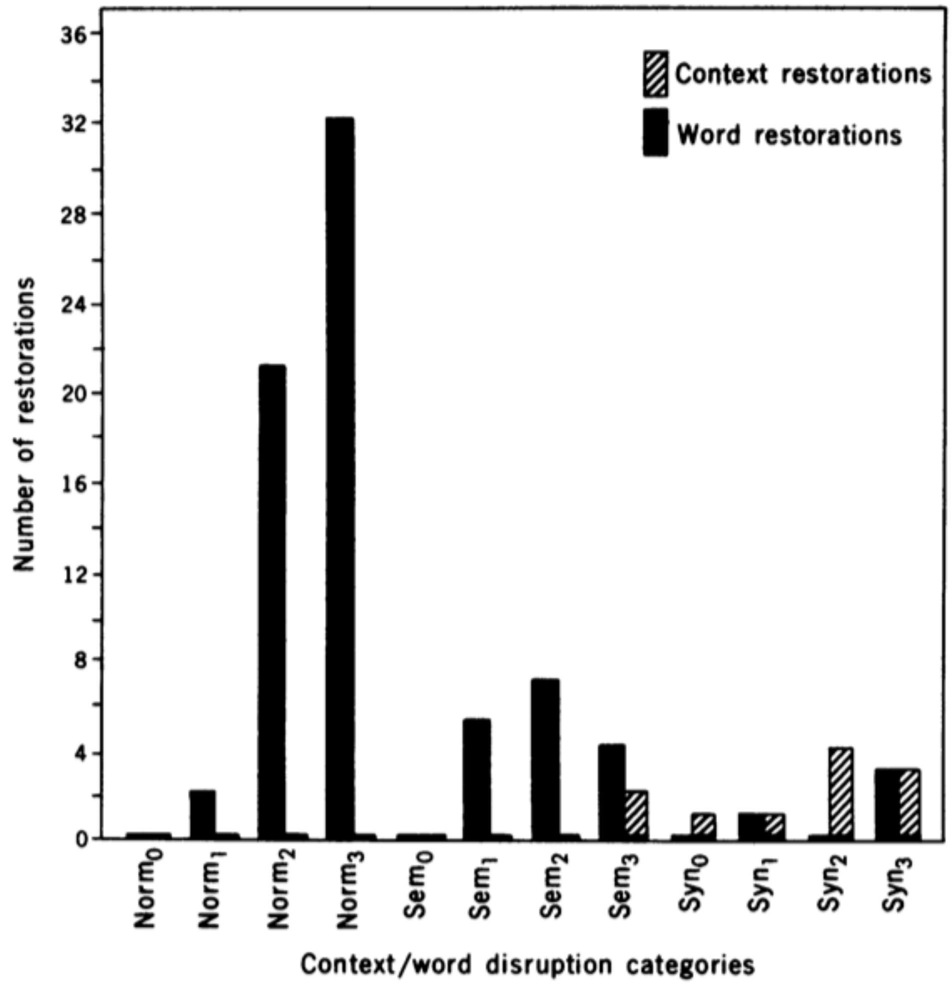
Levels of representation

Semantic	[[touch]]
Syntactic	
Lexical	/tʌtʃ/
Phonetic	

Speech shadowing demo

He thinks he won't get the letter…

Speech shadowing demo



- Linguistic stimuli are analyzed ***incrementally.***

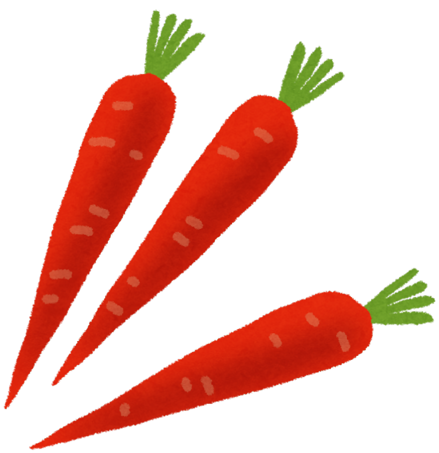
- Analysis occurs at multiple levels in parallel.

Marslen-Wilson 1975

Components of word recognition

Activation / access / generation function:
How are potential word candidates identified, given some sensory input?

PARROT
CARE
CARROT
Kerrianne
Christian



Selection function: How do we make a 'decision' about the outcome of the recognition process?

Integration function: How does output of recognition process relate to higher order levels of analysis (e.g. syntactic or semantic analysis)?

Axes of investigation

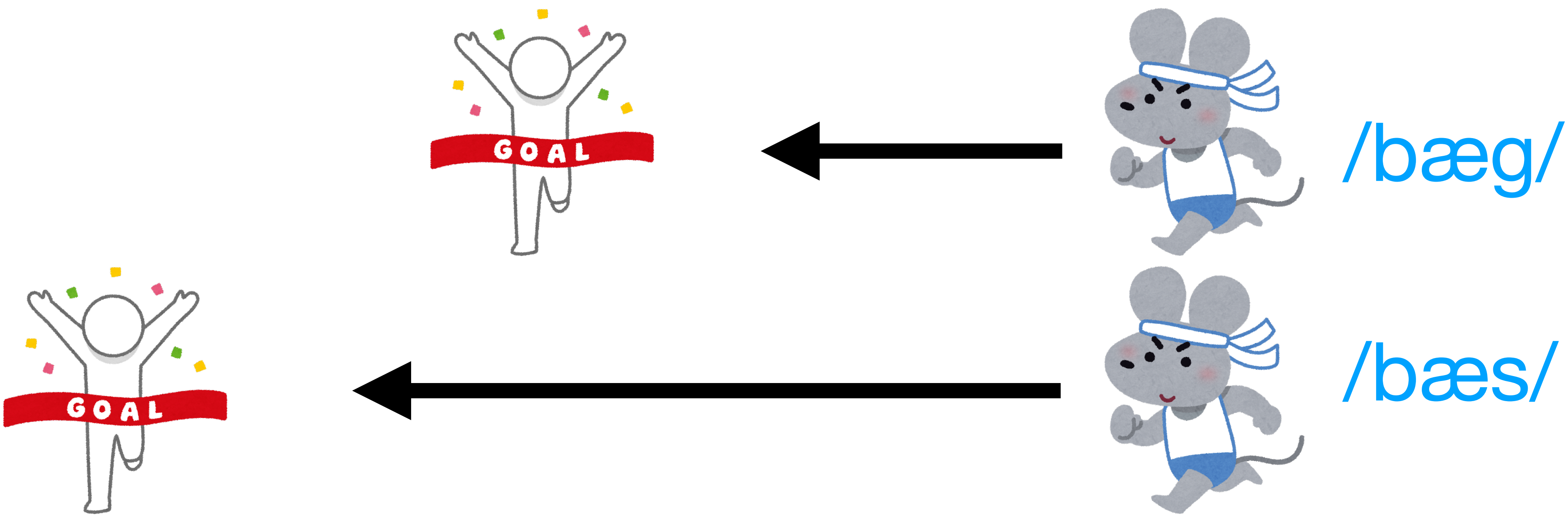
To what degree do these stages interact? Are there isolated cognitive 'modules' that carry out computations at each level of representation ? Or are these different functions composed into a single processing mechanism?

How does information flow in the system? Is the system strictly feedforward (bottom-up), or can higher-order information influence processing at earlier stages (top-down)? If top-down information can influence later processing, how 'far down' can it go?

How are lexical candidates activated in response to sensory input? Are potential candidates activated one by one (serially), or can multiple candidates be activated at once (parallel activation)? Are representations activated by searching for information serially (as in a card catalog) or via direct access (as in a keyword search)?

Logogen model (Morton, 1969)

- Words are activated in parallel in response to speech input.
- A *logogen* is an individual evidence accumulation unit associated with a particular lexical item.
- When a logger reaches its activation threshold, we say that the lexical item has been accessed.
- Words may differ in their activation thresholds: more frequent words might have lower thresholds.

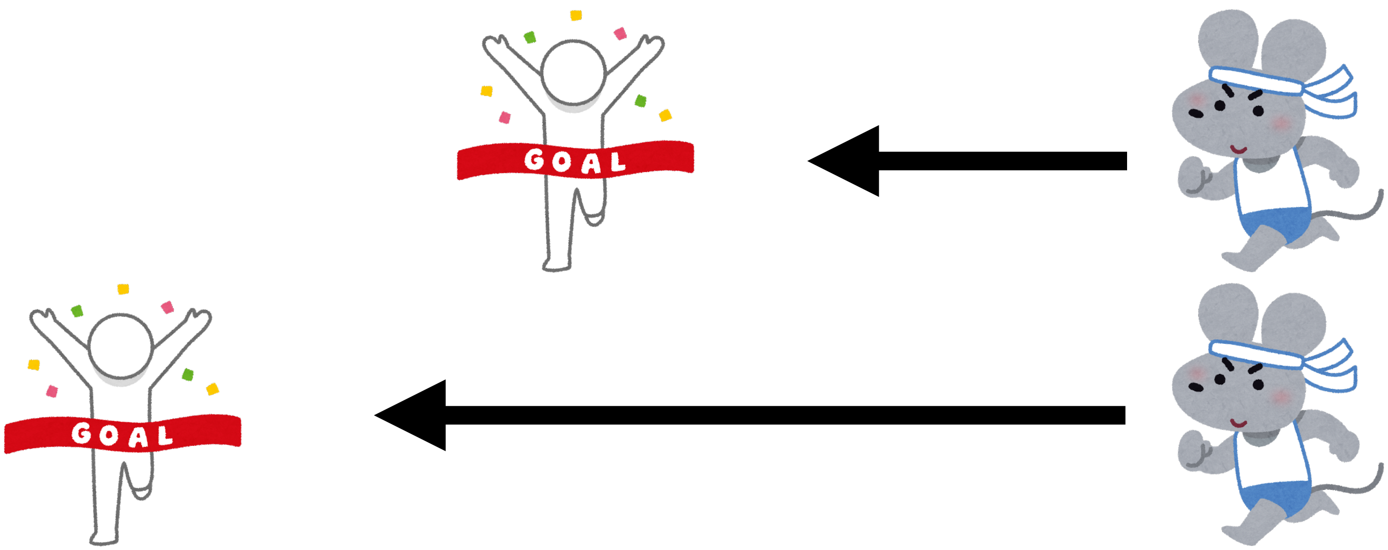


Cohort model (Marslen-Wilson, 1987)

- All possible matching candidates are activated in parallel as the word unfolds. The set of matching candidates is called the **cohort.**
- Candidates are eliminated from consideration once they no longer match the input.
- Candidates are generated bottom-up (i.e. from the speech stream alone), but higher levels of representation may select candidates from cohort.

/k/	→	/kʌ/	→	/kʌp/	→	/kʌp.l/
cat		cut		cup		couple
catch		cup		couple		
candy		cuddle		cupping		
candle		cull		…		
…		…				

Predictions?



- **Speed to recognize non-words?**
- **The time-course of phoneme identification?**
- **Other predictions?**

/k/ → /kʌ/ → /kʌp/ → /kʌp.l/

/k/	/kʌ/	/kʌp/	/kʌp.l/
cat	cut	cup	couple
catch	cup	couple	
candy	cuddle	cupping	
candle	cull		

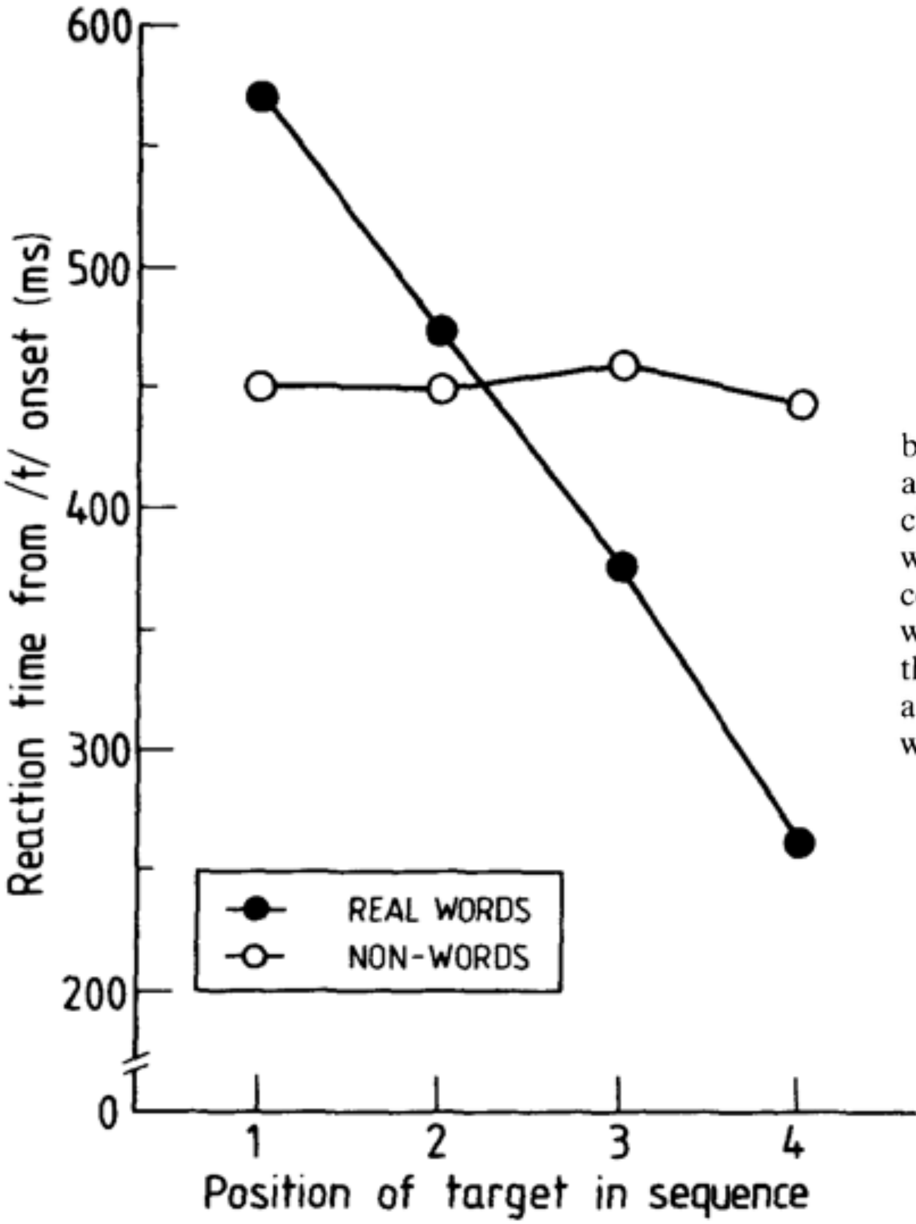
Cohort model (Marslen-Wilson, 1987)

- All possible matching candidates are activated in parallel as the word unfolds. The set of matching candidates is called the **cohort.**
- Candidates are eliminated from consideration once they no longer match the input.
- Candidates are generated bottom-up (i.e. from the speech stream alone), but higher levels of representation may select candidates from cohort.

/k/	→	/kʌ/	→	/kʌp/	→	/kʌp.l/
cat		cut		cup		
catch		cup		couple		
candy		cuddle		cupping		couple
candle		cull		…		
…		…				

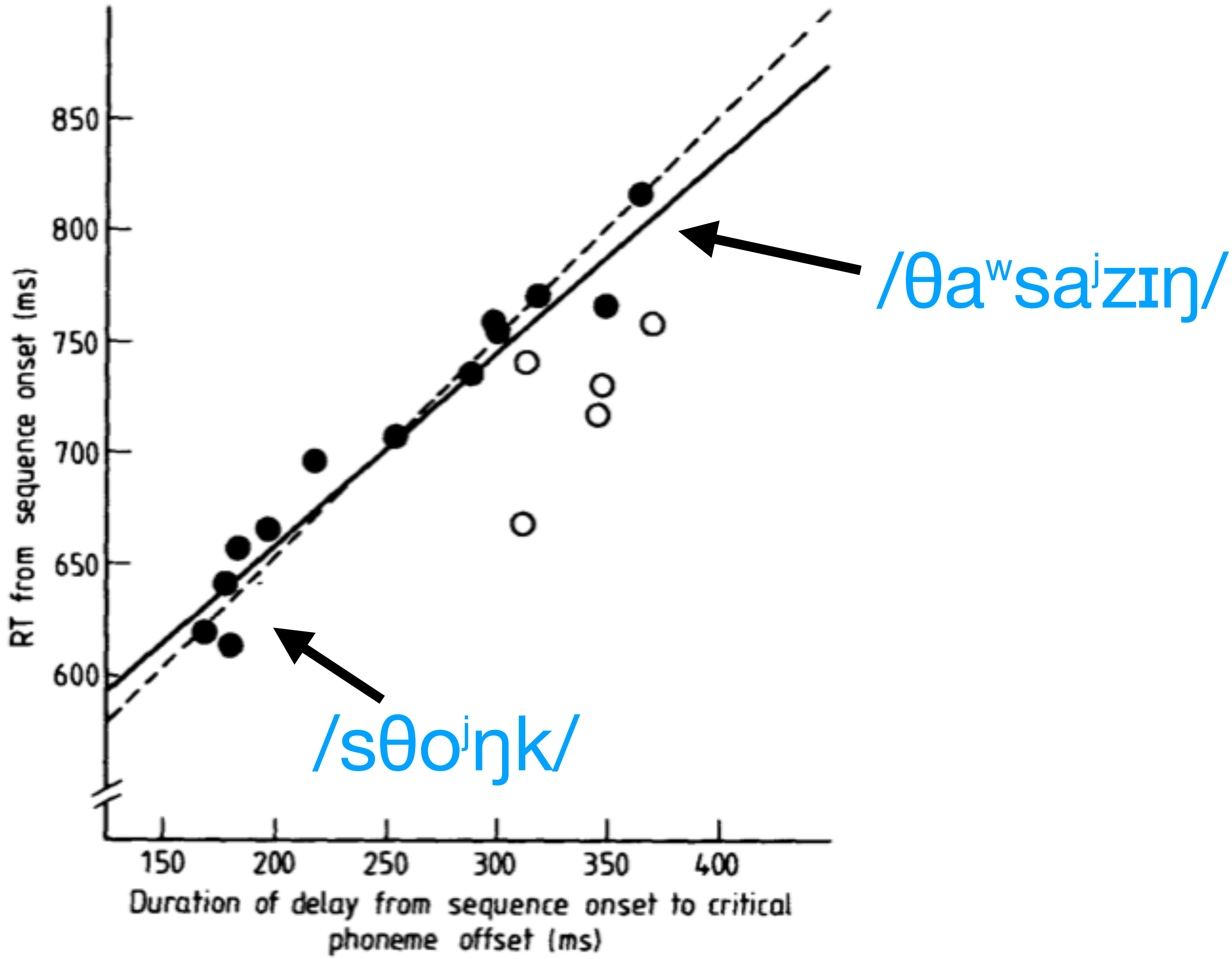
word uniqueness point

Phoneme monitoring (Marslen-Wilson, 1984)



A linear regression analysis showed that there was a close relationship between these distances and the monitoring response ($r = +.89$).[8] The variations in distance accounted for over 80% of the variance in the mean latencies for the 60 individual words containing targets. This strong correlation with phoneme-monitoring latency shows that recognition-points derived from cohort analysis have a real status in the immediate, on-line processing of the word. The subjects in this experiment were using a lexical strategy, so that their response-latencies reflected the timing of word-recognition processes, and the cohort model correctly specified the timing of these processes for the words involved.

Non-word judgments (Marslen-Wilson, 1984)



Cohort model (Marslen-Wilson, 1987)

- All possible matching candidates are activated in parallel as the word unfolds. The set of matching candidates is called the **cohort.**
- Candidates are eliminated from consideration once they no longer match the input.
- Candidates are generated bottom-up (i.e. from the speech stream alone), but higher levels of representation may select candidates from cohort.

/k/	/kʌ/	/kʌp/	/kʌp.l/
cat	cut	cup	couple
catch	cup	couple	
candy	cuddle	cupping	
candle	cull	…	
…	…		

word uniqueness point

Axes of investigation

To what degree do these stages interact? Are there isolated cognitive 'modules' that carry out computations at each level of representation ? Or are these different functions composed into a single processing mechanism?

How does information flow in the system? Is the system strictly feedforward (bottom-up), or can higher-order information influence processing at earlier stages (top-down)? If top-down information can influence later processing, how 'far down' can it go?

How are lexical candidates activated in response to sensory input? Are potential candidates activated one by one (serially), or can multiple candidates be activated at once (parallel activation)? Are representations activated by searching for information serially (as in a card catalog) or via direct access (as in a keyword search)?

Swinney (1979)

→ **Cross modal priming task:** One stimulus is presented auditorily, while a lexical decision task interrupts at a certain point. If the context activates associated meanings at the point of the decision, it should prime (speed up) the lexical decision.

TABLE 1

SCHEMATIZED SAMPLE OF EXPERIMENTAL MATERIALS

	Ambiguity condition	
Context condition	Ambiguous	Unambiguous
No context	Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several bugs$_{\Delta}$ in the corner of his, room.	Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several insects$_{\Delta}$ in the corner of his room.
Biasing context	Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and other bugs$_{\Delta}$ in the corner of his room.	Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and other insects$_{\Delta}$ in the corner of his room.

Visual words Displayed at "Δ"	ANT	(contextually related)
	SPY	(contextually inappropriate)
	SEW	(unrelated)

Swinney (1979)

→ **At short SOAs (stimulus onset asynchronies):** Priming was observed for both contextually related (ANT) and contextually inappropriate words (SPY). This was not seen for unambiguous words.

MEAN REACTION TIMES, IN MILLISECONDS, FOR CONDITIONS OF THE AMBIGUITY X CONTEXT X VISUAL WORD INTERACTION: EXPERIMENT 2 (REPLICATION)

Prime'd!

Ambiguity condition	Context condition	Visually presented word: Contextually related	Contextually inappropriate	Unrelated
Ambiguous	Biasing context	708	715	746
	No context	703	708	743
Unambiguous	Biasing context	710	747	744
	No context	702	732	742

Swinney (1979)

→ **At long SOAs (stimulus onset asynchronies):** Priming was observed only for contextually related (ANT) words, for both ambiguous and unambiguous words.

TABLE 4

MEAN REACTION TIMES, IN MILLISECONDS, FOR CONDITIONS OF THE AMBIGUITY X CONTEXT X VISUAL WORD INTERACTION: EXPERIMENT 2 (EXTENSION)

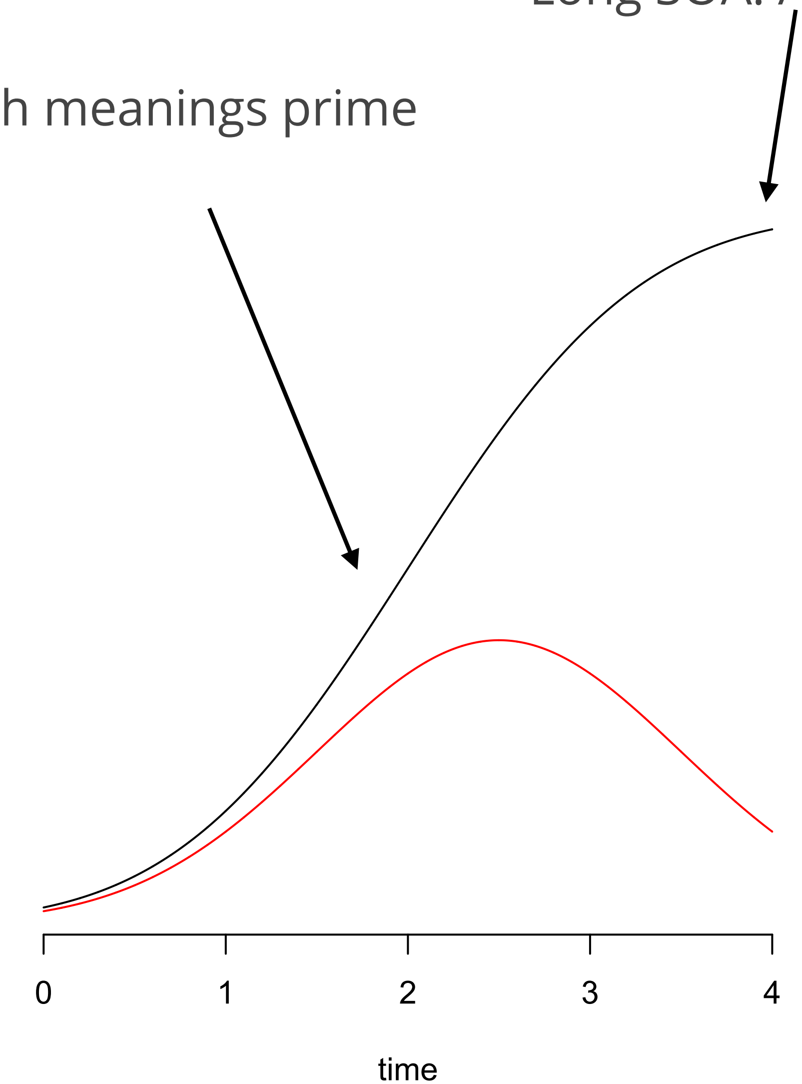
Prime'd!

		Visual word condition		
Ambiguity condition	Context condition	Contextually related	Contextually inappropriate	unrelated
Ambiguous	Biasing context	795	849	848
	No context	800	846	845
Unambiguous	Biasing context	808	843	849
	No context	811	847	846

Swinney (1979)

Long SOA: Appropriate meaning primes

Short SOA: Both meanings prime



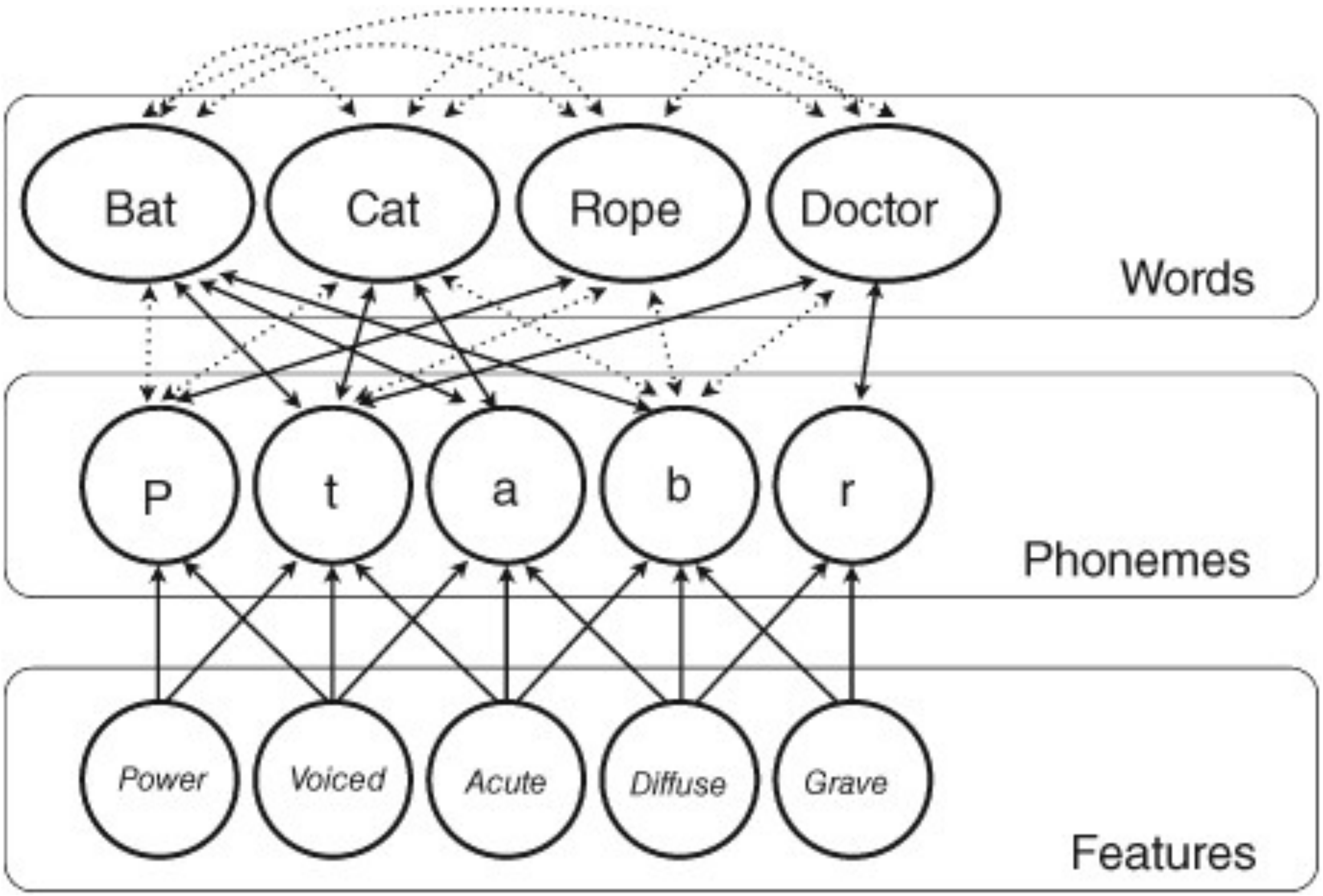
Axes of investigation

To what degree do these stages interact? Are there isolated cognitive 'modules' that carry out computations at each level of representation ? Or are these different functions composed into a single processing mechanism?

How does information flow in the system? Is the system strictly feedforward (bottom-up), or can higher-order information influence processing at earlier stages (top-down)? If top-down information can influence later processing, how 'far down' can it go?

How are lexical candidates activated in response to sensory input? Are potential candidates activated one by one (serially), or can multiple candidates be activated at once (parallel activation)? Are representations activated by searching for information serially (as in a card catalog) or via direct access (as in a keyword search)?

TRACE



→ **TRACE** model: Lexical processing involves top-down feedback from activated guesses about the lexical item being processed.

TRACE

→ **Visual world paradigm:** Experimental paradigm where participants are given instructions on how to interact with a display (e.g. *please click on the picture of a beetle*), while their eye movements are continuously tracked.

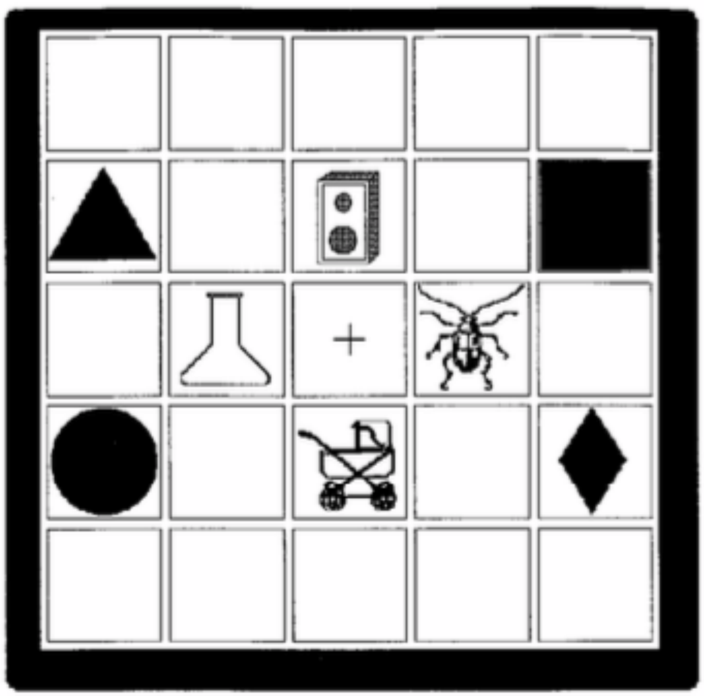


FIG. 3. An example of a stimulus display presented to participants.

Allopenna et al., 1998

TRACE

Computer simulations with TRACE **Human data**

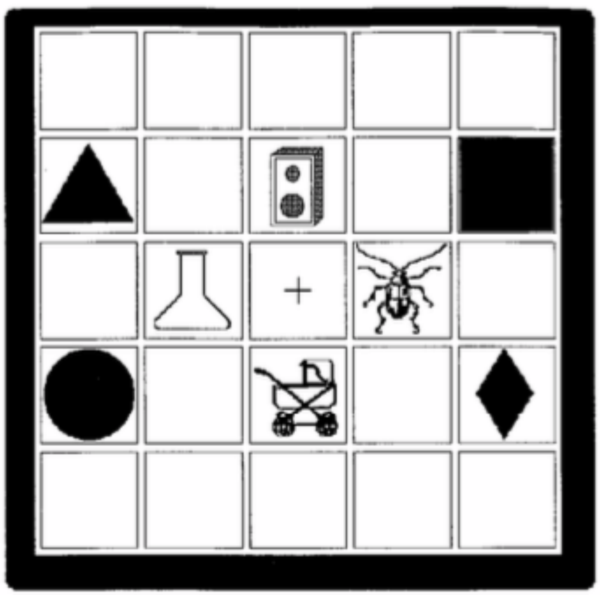


FIG. 3. An example of a stimulus display presented to participants.

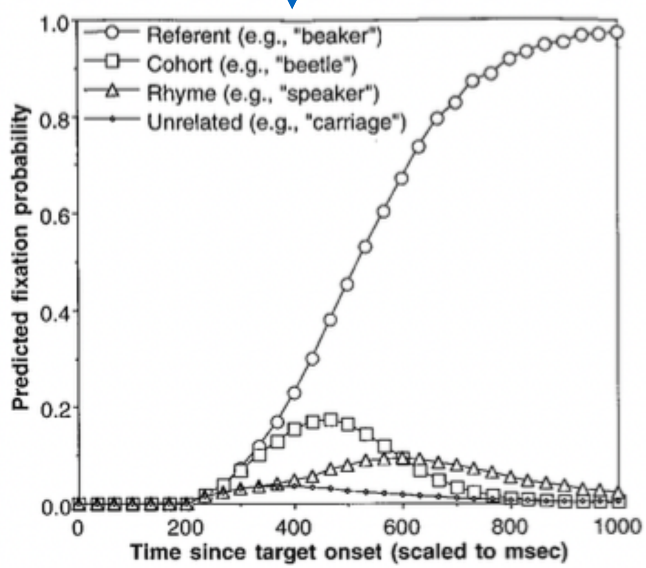


FIG. 2. Predicted response probabilities converted from TRACE using the scaled Luce choice rule.

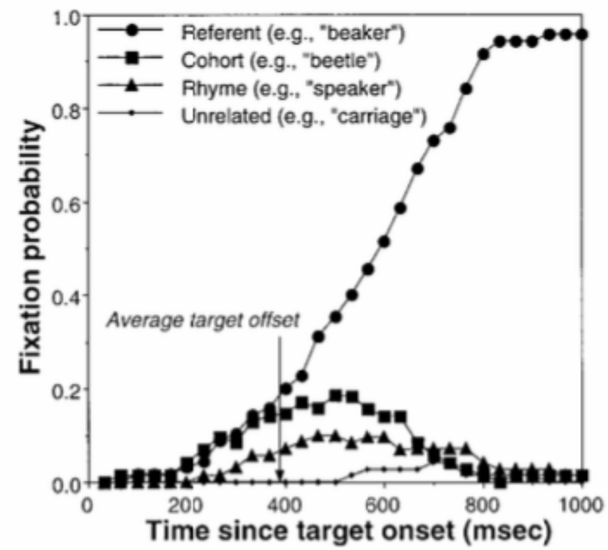


FIG. 4. Probability of fixating each item type over time in the full competitor condition in Experiment 1. The data are averaged over all stimulus sets given in Table 1; the words given in the figure are examples of one set.

Allopenna et al., 1998

Axes of investigation

To what degree do these stages interact? Are there isolated cognitive 'modules' that carry out computations at each level of representation ? Or are these different functions composed into a single processing mechanism?

How does information flow in the system? Is the system strictly feedforward (bottom-up), or can higher-order information influence processing at earlier stages (top-down)? If top-down information can influence later processing, how 'far down' can it go?

How are lexical candidates activated in response to sensory input? Are potential candidates activated one by one (serially), or can multiple candidates be activated at once (parallel activation)? Are representations activated by searching for information serially (as in a card catalog) or via direct access (as in a keyword search)?