Cognition

PSYC 2040

L6: Information Processing

Part 2

logistics: class survey (February)

- https://forms.gle/NyGRXCy6grEgiVPm9
- link also on Canvas (under class surveys)
- due Feb 27 (**Tues midnight**, so we can talk about it in class on Wed)
- 0.5 extra credit point that counts towards your final points/grade
  - submit on Canvas (it’s an “assignment” on Canvas)
- I value your feedback
- anonymous survey! please be honest and reflective
- you will get a code at the end of the survey (on the thank you screen)
  - copy-paste this code on Canvas to get credit

logistics: midterm + monthly quiz

monthly quiz

- available from Friday (Feb 23) to Tuesday (Feb 27) midnight
- open-book, Canvas
- 1 hour time limit

practice assessment 1

- multiple-choice + short answers
- available on Canvas
- will post answer keys next week

review sessions

- Monday (Feb 26), 7-9 pm
- Thursday (Feb 29), 8-10 pm
- Kanbar 200

midterm

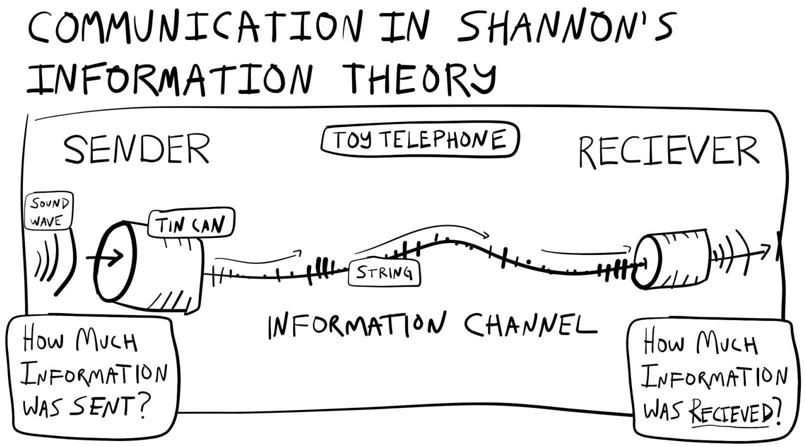
- March 1
- in-person
- Canvas quiz + handwritten short answer
- closed-book

recap

- what we covered:
  - metaphors for cognition
  - Donders' subtractive logic
- your to-dos were:
  - *do:* PRP experiment
  - *explore*: L6 assignments

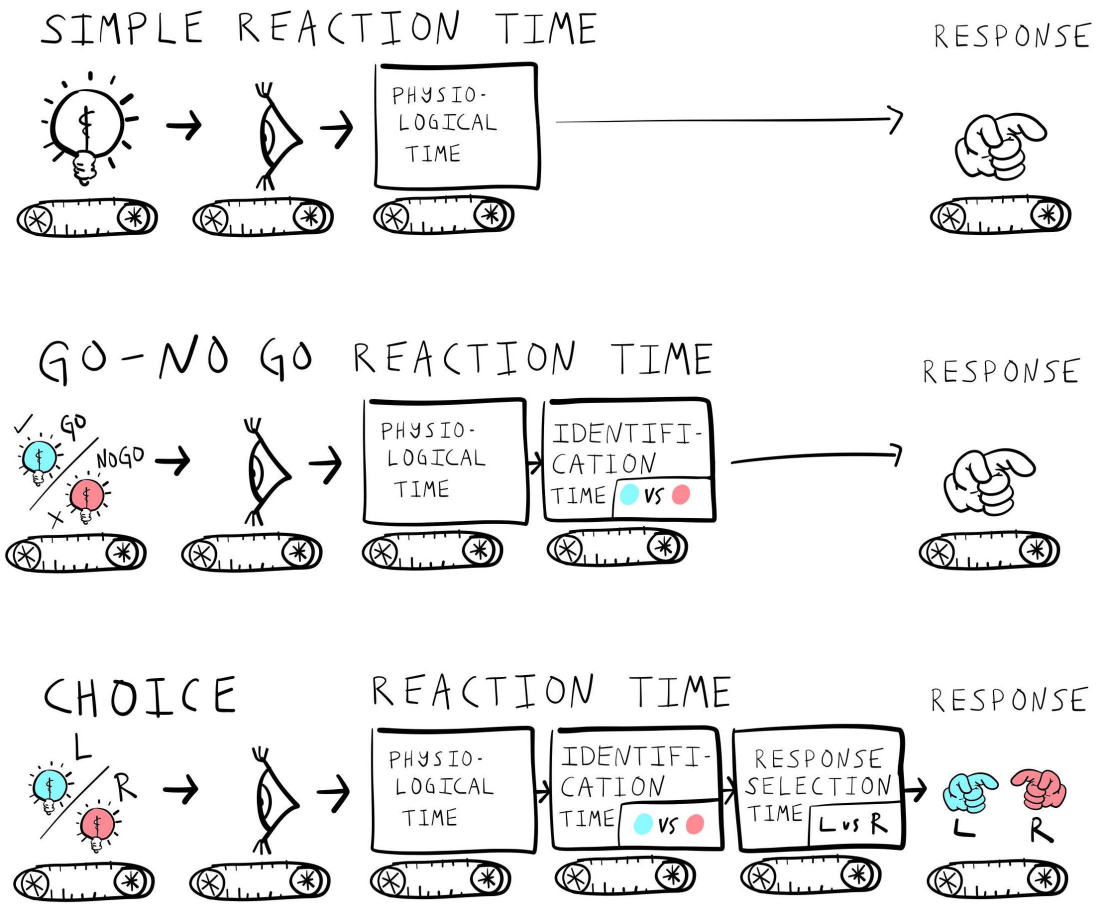
today’s agenda

- PRP effect
- Shannon’s information theory
- the telephone metaphor for cognition
- from behaviorism to cognitivism



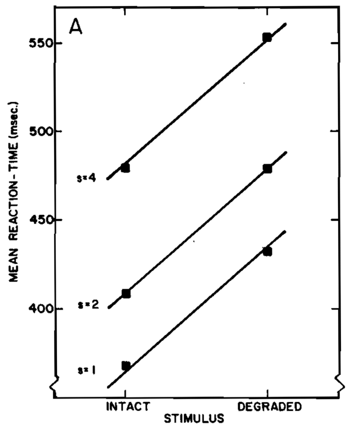
recap: Donders' subtractive logic

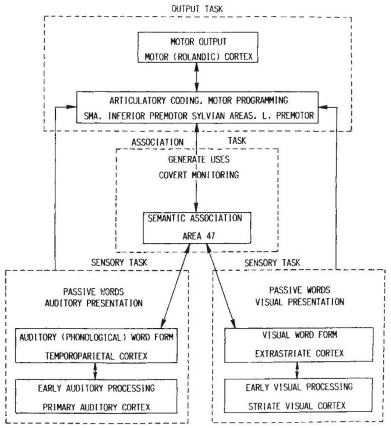
- time taken to respond should depend on number of processing stages required to complete the task
  - simple tasks have fewer stages and are therefore performed quickly
  - complex tasks have more stages and therefore performed slower



recap: additive vs. interactive factors

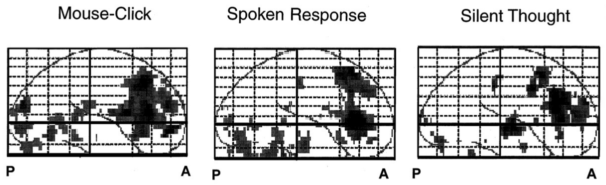
- Sternberg (1969): binary classification RT studies
  - additive effects suggest that two variables do not interactively influence the dependent variable
- Peterson et al. (1998): PET study, verb use task
  - difference of brain images helped isolate specific areas for specific cognitive processing
- Jennings et al. (1997): PET study, semantic vs. letter task across three response modalities
  - behavior showed no interactions, neural responses showed interaction between task and modality
- key takeaways:
  - cognitive signatures ≠ neural signatures
  - subtractive logic may have its limits (insertion assumption)





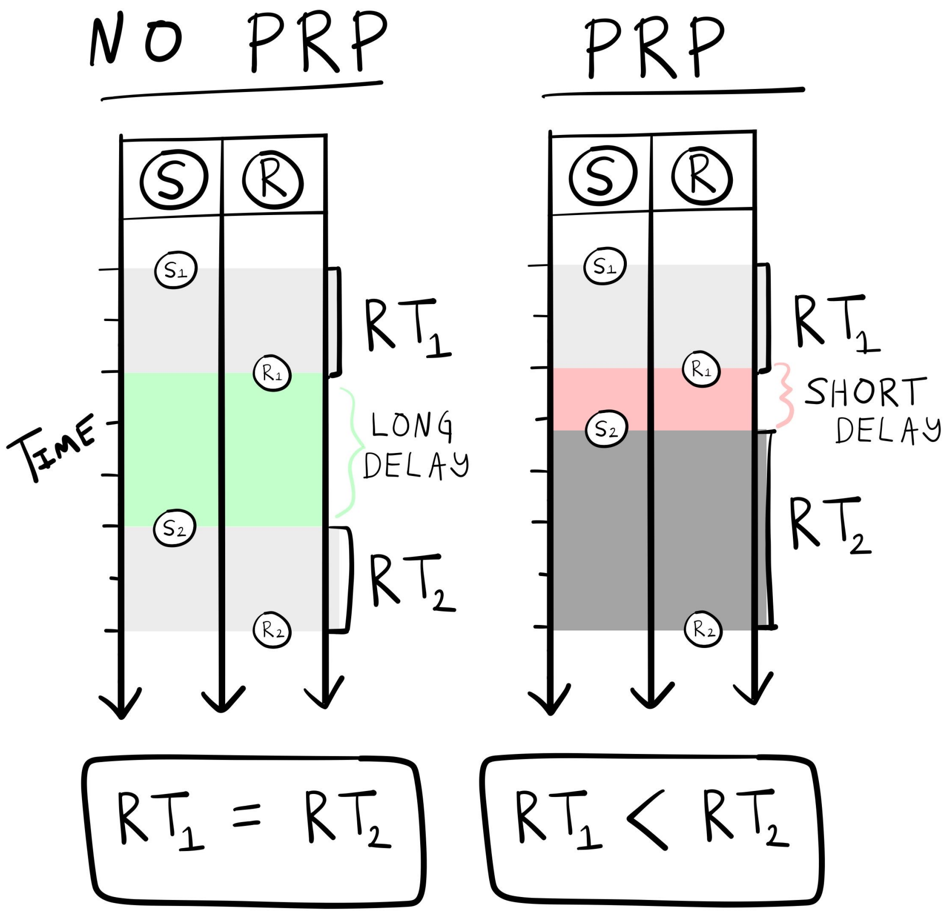
Probability of Responding “Old” to Old and New Items on the Recognition Test for Each Response Mode

	Mouse-click		Spoken response		Silent thought	
Item	Semantic	Letter	Semantic	Letter	Semantic	Letter
Old	0.90	0.52	0.87	0.52	0.79	0.54
New	0.27	0.30	0.22	0.24	0.28	0.23



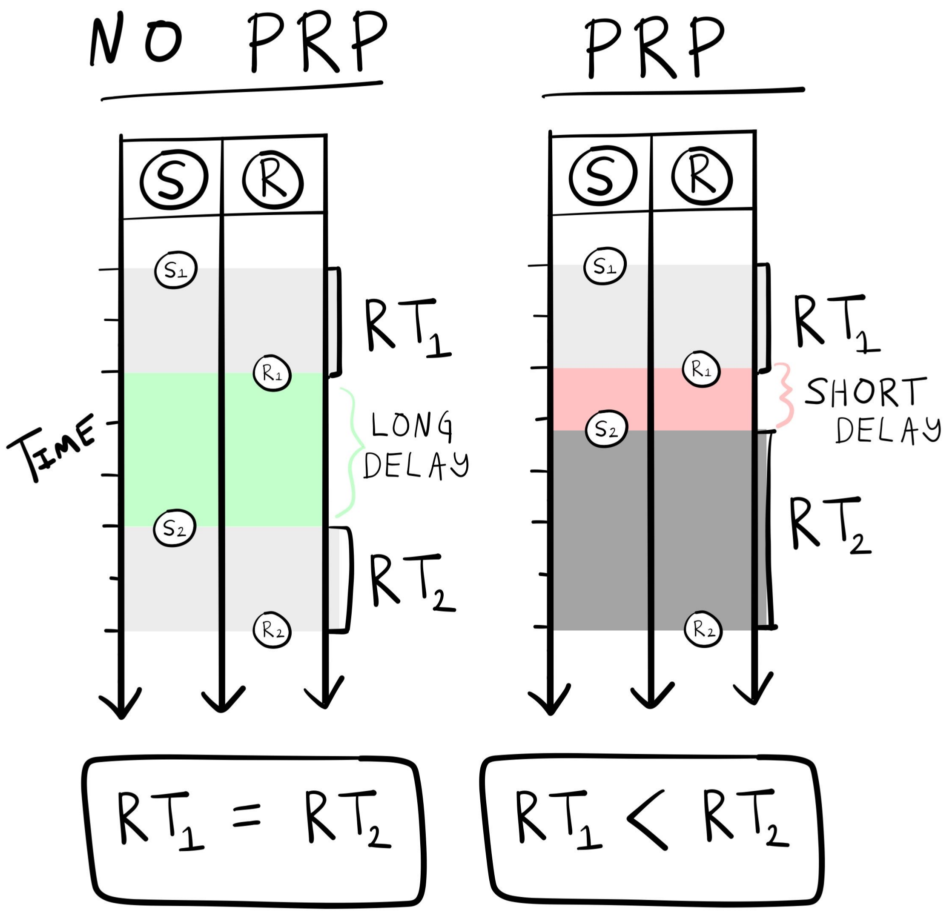
PRP effect

- the psychological refractory period (PRP) effect was documented by A.T. Welford
- the idea was that if two identical stimuli (S1 and S2) are presented with a short delay, then the time taken to respond to S2 is longer (RT2 > RT1)



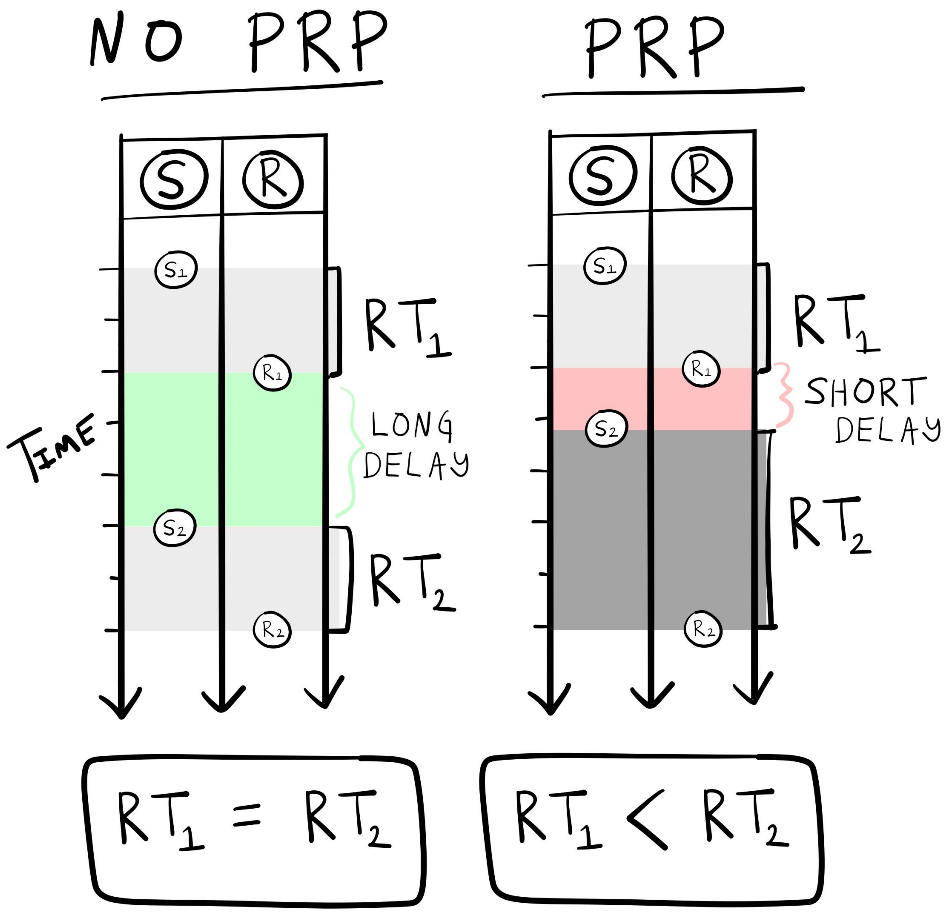
PRP effect: real-life examples

- groups of 2
- come up with a real-life example
- debrief

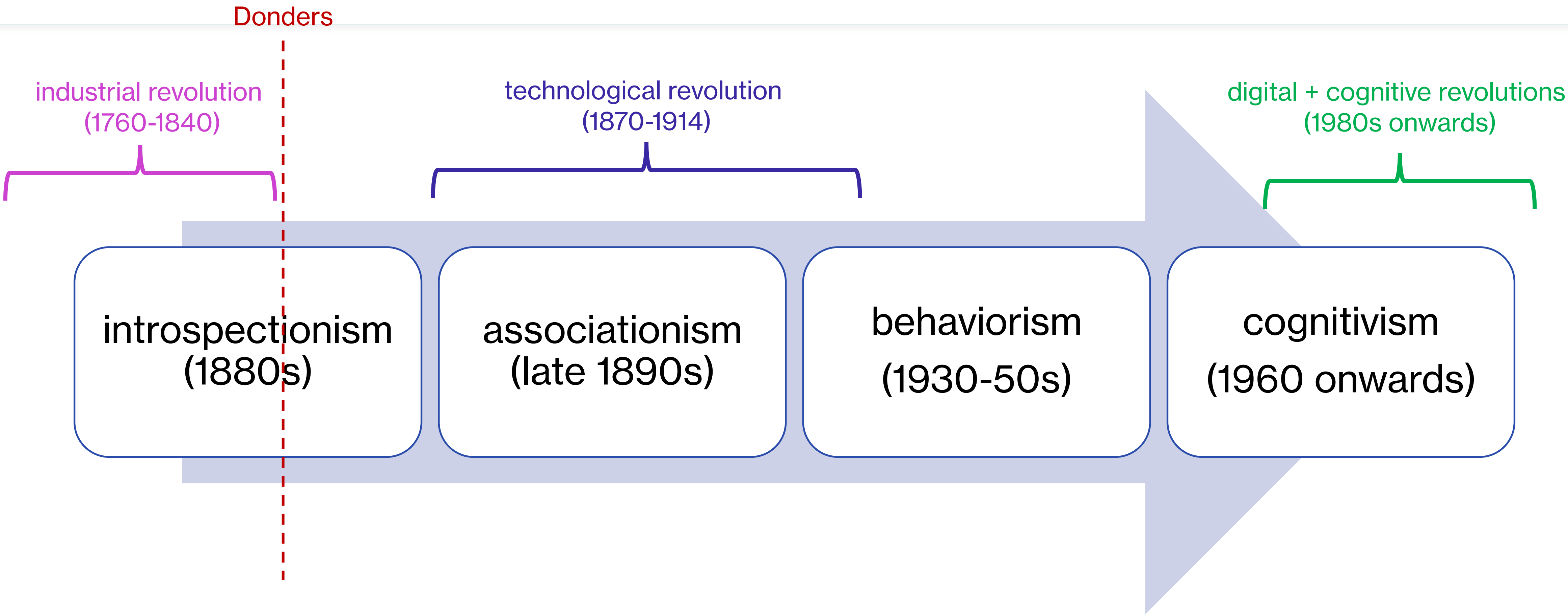


PRP effect: explanations

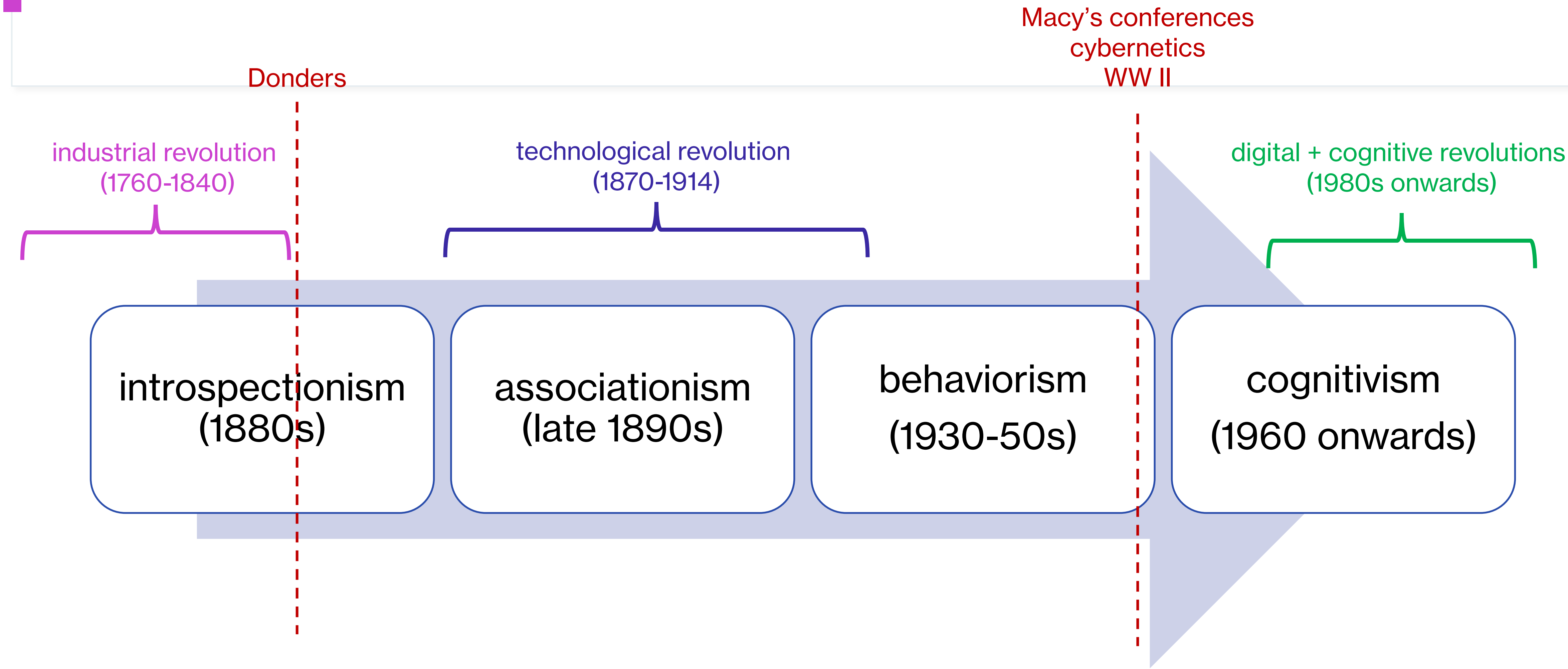
- properties of nerve fibers
- participant surprise: shorter delays produce more surprise which increases time
- limited-capacity single channel
    - inspired by the assembly line metaphor and how a bottleneck might be created if stimuli were presented quickly one after the other
    - also inspired by telecommunications…the idea of a “single channel”



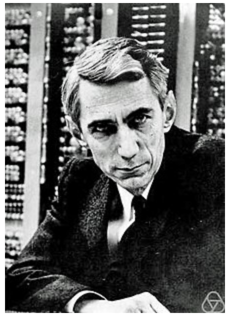
the timeline so far



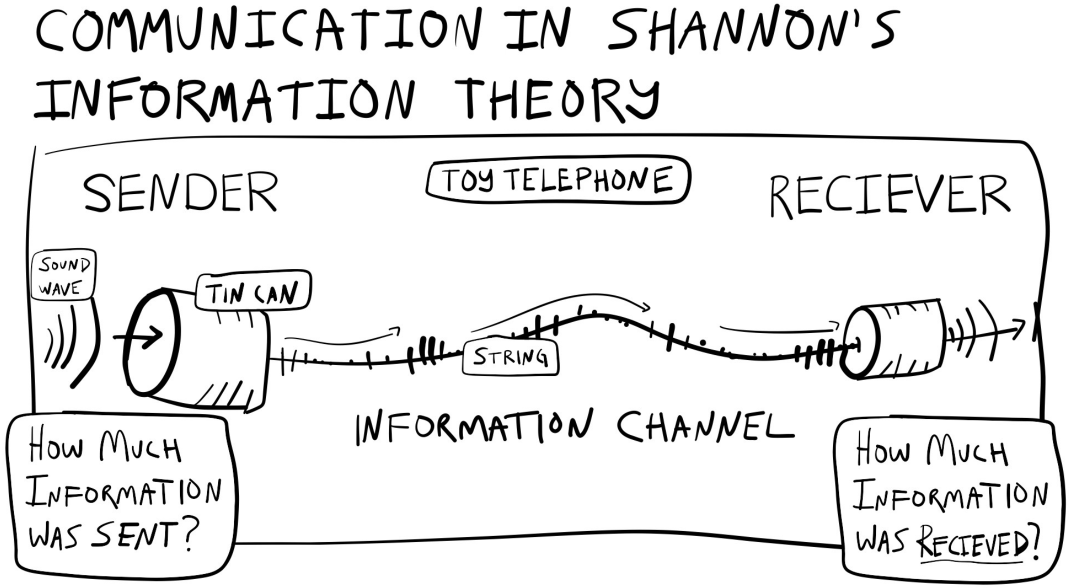
the timeline so far



origins: information theory

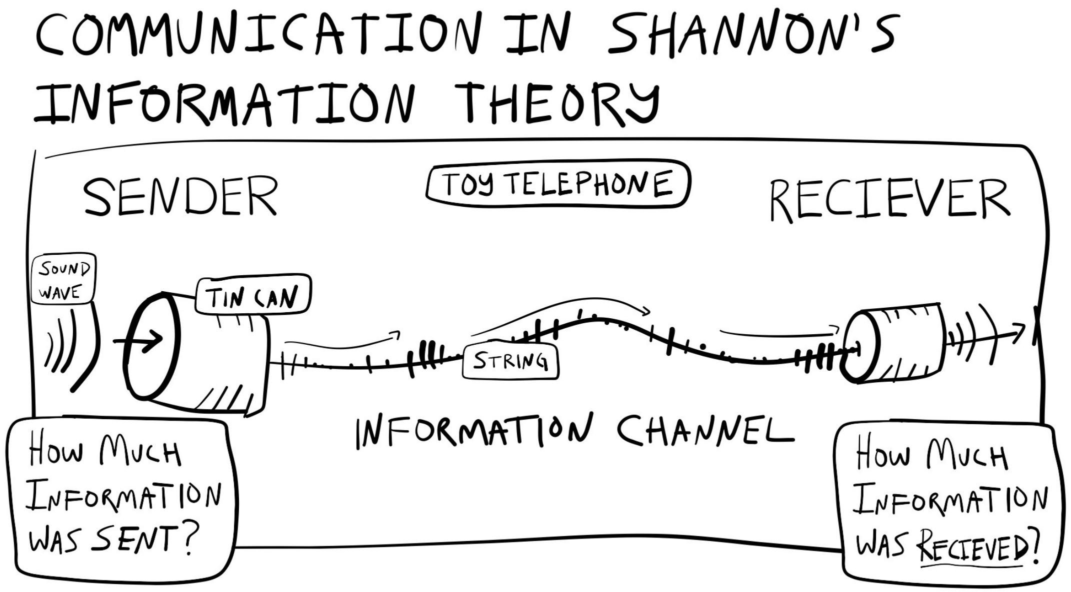


- Claude Shannon, American mathematician regarded the "father of information theory"
  - also contributed to cryptanalysis during WWII
- the main purpose of information theory was to characterize communication systems, not understand cognition/psychology



information channels

- three main components
  - sender
  - channel
  - receiver
- channels have capacities
  - the total amount of information that can be transmitted
  - how many calls can one telephone line/string support?
  - how many stimuli can be processed within a given time period?



measuring channel capacity

- Shannon proposed a mathematical formula for quantifying the amount of information via *H*, or entropy, i.e., the amount of uncertainty/randomness/noise in a system of messages
- key idea: the more predictable something is, the less information it can transmit
  - a book of all As provides no new information, i.e., its entropy (H) could be 0
  - a 5-sentence paragraph with many new concepts and combinations of words has high entropy, i.e., more information to transmit

$$H(X) = -1 * \sum_{\mathrm{i}=1}^{n} P(x_i) * log_2 P(x_i)$$

H : entropy / information

$P(x_i)$: probability of occurrence for each event x_i

$log_2 P(x_i)$: log of same probability

*: multiply

example 1: measuring channel capacity

- consider a fair coin, we want to calculate how much “information” it can transmit
- there are two events
  - x_1 = heads and x_2 = tails
  - $P(x_1) = P(x_2) = 0.5$
  - $\log(P(x_1)) = \log(P(x_2)) = -1$
  - $H(X) = -1 * \text{sum}(0.5(-1) + 0.5(-1)) = -1 * -1$
  - $H(X) = 1$

$$H(X) = -1 * \sum_{\mathrm{i}=1}^{n} P(x_i) * log_2 P(x_i)$$

H : entropy / information

$P(x_i)$: probability of occurrence for each event x_i

$\log_2 P(x_i)$: log of same probability

*: multiply

example 2: measuring channel capacity

- now consider an unfair coin, we want to calculate how much “information” it can transmit
- there are two events
  - x_1 = heads and x_2 = tails
  - $P(x_1) = 0.8$, $P(x_2) = 0.2$
  - $\log(P(x_1)) = -0.32$, $\log(P(x_2)) = -2.32$
  - $H(X) = -1 * \text{sum}\,(0.8(-.32) + 0.2(-2.32))$
  - $H(X) = -1 * -0.72 = 0.72$
- an unfair coin is less random than a fair coin and therefore has lower “information” to transmit, i.e., lower entropy

$$H(X) = -1 * \sum_{\mathrm{i}=1}^{n} P(x_i) * log_2 P(x_i)$$

H : entropy / information

$P(x_i)$: probability of occurrence for each event x_i

$\log_2 P(x_i)$: log of same probability

*: multiply

activity: measuring channel capacity

- calculate the entropy of a dice
- groups 1-3
  - a fair dice
  - all $P(x_i) = 0.167$ for all numbers
- groups 4-6
  - an unfair dice
  - $P(x_1) = 0.90$ for 1
  - $P(x_i) = 0.02$ for all other numbers

$$H(X) = -1 * \sum_{\mathrm{i}=1}^{n} P(x_i) * log_2 P(x_i)$$

H : entropy / information

$P(x_i)$: probability of occurrence for each event x_i

$log_2P(x_i)$: log of same probability

*: multiply

bits of information

- H uses a base 2 logarithm to produce a number in the unit of bits
- bits refer to the total number of discrete events in a system of messages, **it is a unit of information**
- one bit has two states: 0 or 1
  - it could be used to represent two events/states
  - e.g., heads or tails, on or off
- 2 bits can be of the form 00, 01, 10, 11
  - 4 events could be represented by 2 bits
- general formula
  - number of events = 2^{bits}

# OF BITS	COMBINATIONS	# OF EVENTS
1	0, 1	2
2	1 00, 2 01, 3 10, 4 11	4
3	1 000, 2 001, 3 010, 4 011, 5 100, 6 101, 7 110, 8 111	8

2^{BITS} = # OF UNIQUE EVENTS

$2^1 = 2$

$2^2 = (2 \times 2) = 4$

$2^3 = (2 \times 2 \times 2) = (4 \times 2) = 8$

a communication game

- suppose you (sender) had to transmit the outcome of a dice roll to your friend (receiver)
  - your signal could be one of 6 “events”
  - how many bits? (# events = 2^{bits})
  - more than 2 and less than 3 bits
- recall that when you calculated this for a fair and unfair dice
  - H = 2.58 (fair) and H = 0.70 (unfair)
- when predictability is low (fair dice), you need more bits
- when predictability is high (unfair dice), you need fewer bits

bit activity

- let's play a game in groups of 3 (knower, asker, recorder)
  - earliest (knower) to latest (recorder) birthday in the year
  - each group will be presented with a sheet of paper with 4 or 8 squares
- knower:
  - one of the squares will have a star
  - only you know the location of this star (go to this sheet, KNOWERS ONLY!)
- asker
  - you will not know which of the squares contains the star
  - you can ask N yes/no questions to determine where the star is
- recorder
  - record how many yes/no questions were used to determine the star's location
  - record **max** number of yes/no questions that would be required to CERTAINLY know the location

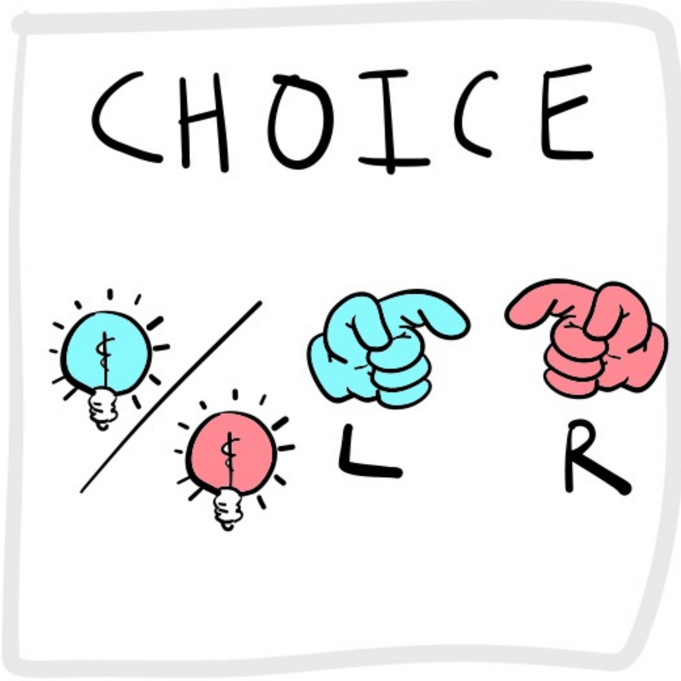
activity inspired from https://www.youtube.com/watch?v=6HCFOyZI9tA

bits activity: debrief

- the “squares” in this game could be considered “events” with equal probability
  - the star is equally likely to be in any one of the squares
  - 1 bit is equivalent to 1 yes/no question
  - #events = $2^{\#bits}$; 4 squares need 2 bits
  - bits represent a lower bound on how many “questions” need to be asked to fully reveal a message
- in communication, we want to know how many bits are needed to convey a particular message because channel capacity (how many bits can be used) is limited
  - if you had a channel capacity of 2 bits, you could only ask 2 yes/no questions
- internet/broadband speeds are encoded in bits!
  - “mbps” stands for megabits per second (1 million bits per second)
  - this refers to the channel capacity, i.e., how many bits can be transmitted in one second
- bits in cognition: “information” contained in a set of stimuli

applying information theory to cognition

- researchers in the 1950s were inspired by the work in telecommunications and applied information theory to the study of cognition
- one domain where these ideas gained prominence was choice reaction time tasks or N-AFCs
  - examples from last class?



choice RTs: set size effects

- one finding from the literature was that the choice reaction time increased as the number of alternatives increased
  - RTs were faster in two vs. four-alternative tasks
  - how many bits to represent two alternatives vs. four alternatives?
- but why? was it the number of alternatives (2 vs 4) or the amount of information (bits) carried within the alternatives (1 vs 2)?
- previous experiments had confounded the number of alternatives and amount of information

Hick Hyman's experiments

- experiment 1
  - choice reaction time task
  - 8 conditions corresponding to different number of alternatives (1 to 8)
  - 1 alternative = 0 bits, 2 alternatives = 1 bit, etc.
  - alternatives were confounded with bits
- experiment 2
  - systematically varied the bits and alternatives
  - how would you design such an experiment?

Condition	Number of alternatives	P (event)	bits = -1 * sum ($P(\log_2(P))$)
1	2	9/10, 1/10	0.47
2	2	8/10, 2/10	0.72
3	4	13/16, 1/16	0.99
4	4	4/8, 3/8, 1/8	

Hick Hyman's experiment 2

TABLE 1

THE EIGHT CONDITIONS FOR EXPERIMENT II AND THE CORRESPONDING AMOUNTS OF INFORMATION IN BITS PER STIMULUS PRESENTATION

Cond.	Number of Alternatives	Probability of Occurrence	$\log_2 1/p$	Av. Amount of Information in Cond.
1	2 { 1 1	9/10 1/10	0.15 3.32	0.47
2	2 { 1 1	8/10 2/10	0.32 2.32	0.72
3	4 { 1 3	13/16 1/16	0.30 4.00	0.99
4	6 { 1 5	15/20 1/20	0.42 4.32	1.39
5	4 { 1 1 2	4/8 2/8 1/8	1.00 2.00 3.00	1.75
6	6 { 1 5	5/10 1/10	1.00 3.32	2.16
7	8 { 1 1 6	8/16 2/16 1/16	1.00 3.00 4.00	2.38
8	8 { 2 2 4	4/16 2/16 1/16	2.00 3.00 4.00	2.75

SAME # OF ALTERNATIVES $\boxed{2}$
DIFFERENT AMOUNT OF BITS

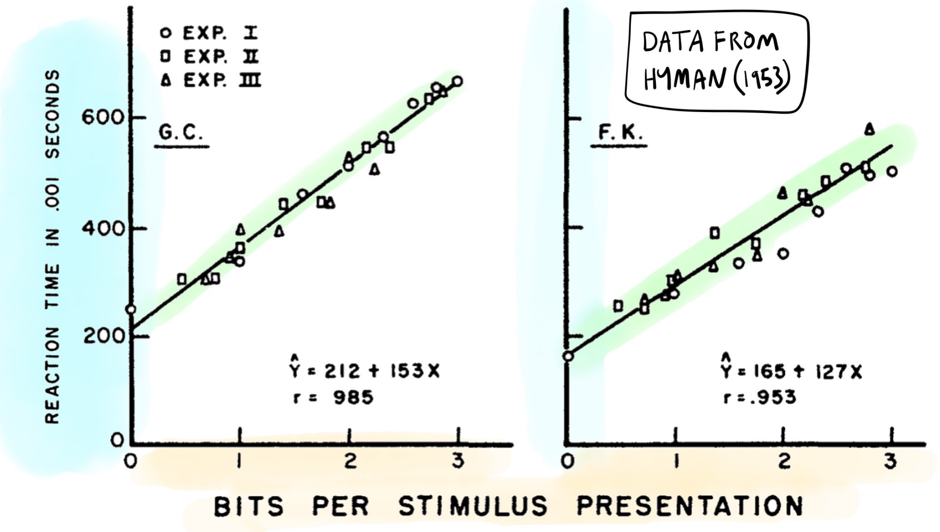
$\boxed{4}$ $\boxed{6}$ $\boxed{8}$

PREDICTION:

REACTION TIME WILL ↑
AS BITS ↑
NOT AS
OF ALTERNATIVES ↑

HICK-HYMAN LAW

CHOICE REACTION TIME INCREASES AS A LINEAR FUNCTION OF THE INFORMATION (BITS) IN THE STIMULUS SET



Hick Hyman's findings: explanations

- match to template hypothesis
  - individuals had "mental templates" of each alternative and were serially comparing the presented stimulus to the templates
  - could not account for the bits/uncertainty of alternatives
- binary logic hypothesis
  - dividing the set of options by half each time
  - popular way to sort numbers in computers (binary sort)
- repetition priming: potential confound
  - fewer alternatives meant more repetitions

Hick Hyman’s findings: broader implications

- debates about interpretation
  - what was the mechanism of how information was processed? Information theory was limited to a measure and did not come with a theory or mechanism
  - violations: practice, set size, etc.
- problematic for behaviorism
  - participants were not simply responding to the stimulus but also thinking about what else could have been presented, i.e., mental operations
  - people started recognizing the value of understanding cognition
  - also highlighted the parallel nature of mental processing
- moving to newer metaphors
  - cognition = computer (Alan Turing & others)

big takeaways

- the study of cognition has moved from introspectionism to associationism to behaviorism to “cognitivism”
- cognition was influenced by world events
- Donders’ processing stages are an example of the assembly line metaphor, inspired from the industrial revolution
- Shannon’s information theory explored the telephone metaphor via the Hick Hyman law for choice reaction times
- 1940-50s onwards was an active period where behaviorism was powerful over time, the value of exploring internal mental operations was recognized

next class

- **before** class:
  - *finish:* L6 quiz + writing assignments
  - *complete:* monthly quiz
  - *review*: practice assessment 1
  - *fill out*: feedback survey
- **during** class:
  - review L0-L6, bring questions!