Cognition: Methods and Models

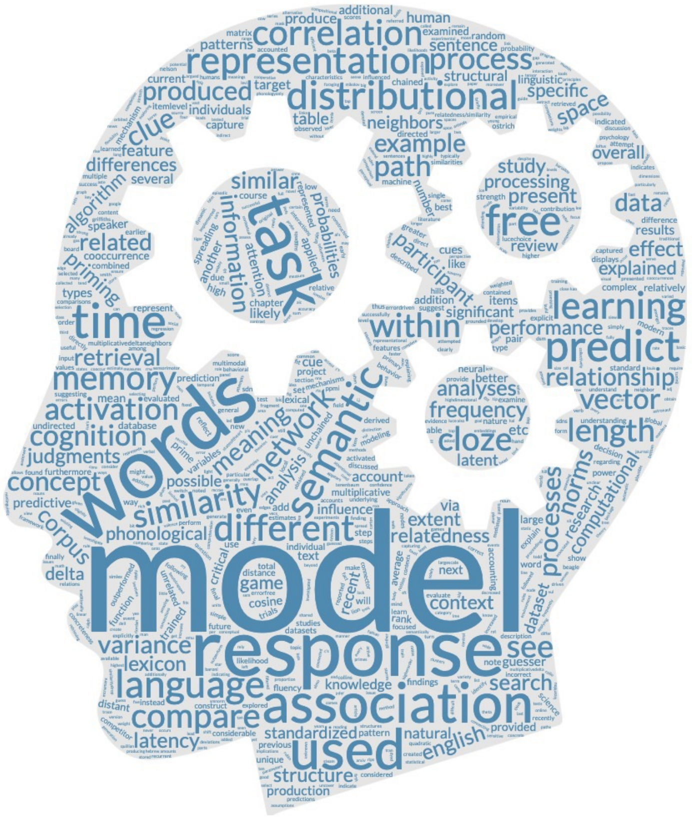
PSYC 2040

L6: Information Processing

Part 2

logistics: research talk today!

- stop by if you're interested in interdisciplinary research and the [lexicon lab]
- 5 pm, Kanbar 107
- agenda:
  - 20-minute overview of research questions + projects
  - pathways to get involved
  - skills you can acquire
  - questions you may have



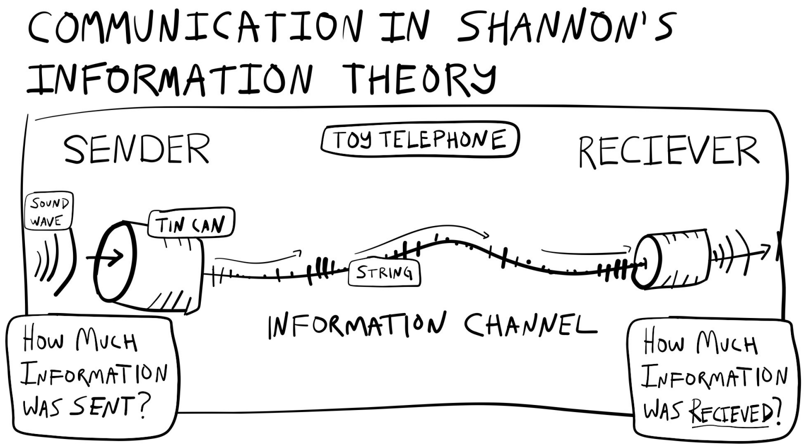
canvas walkthrough: milestones

recap: Feb 28, 2023

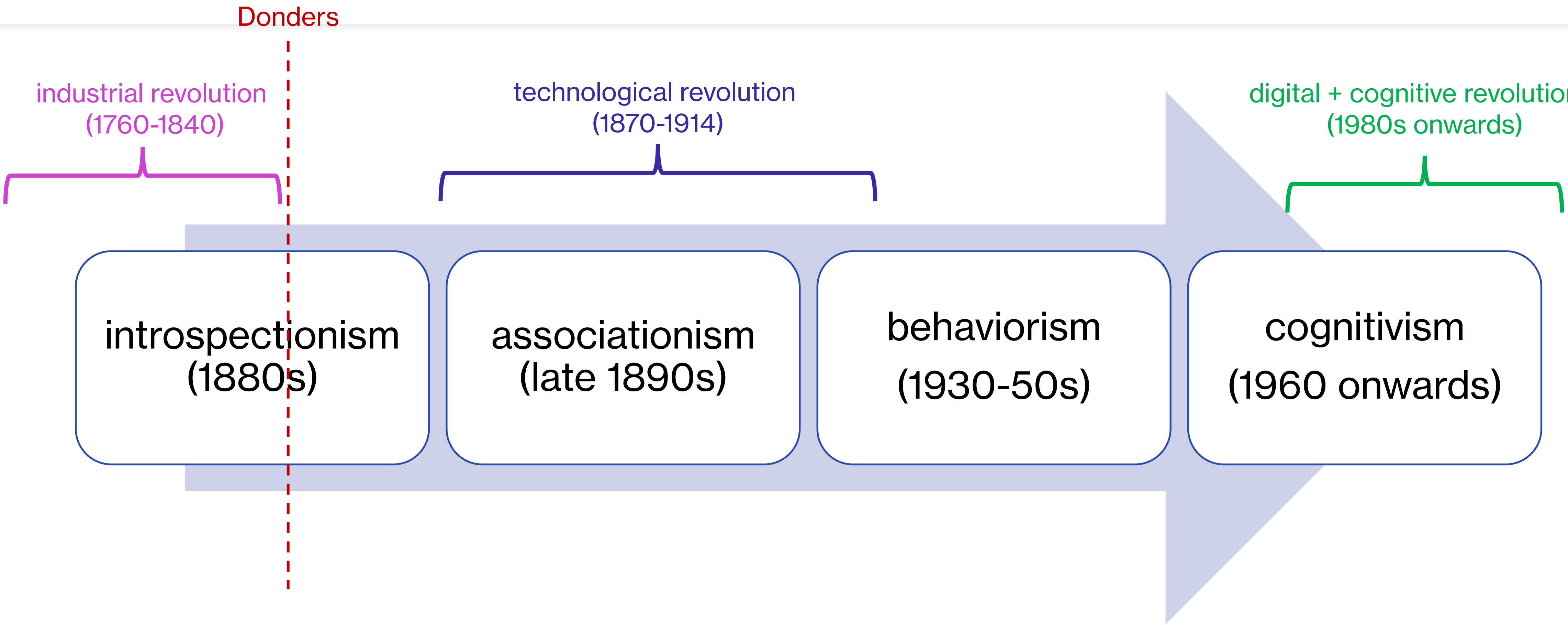
- what we covered:
  - metaphors for cognition
  - Donders + PRP effect
- your to-dos were:
  - *work on:* project milestone #3
  - *post*: conceptual question
  - *block out time*: practice assessment 1
  - *explore*: L6 assignments

today’s agenda

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



the timeline so far



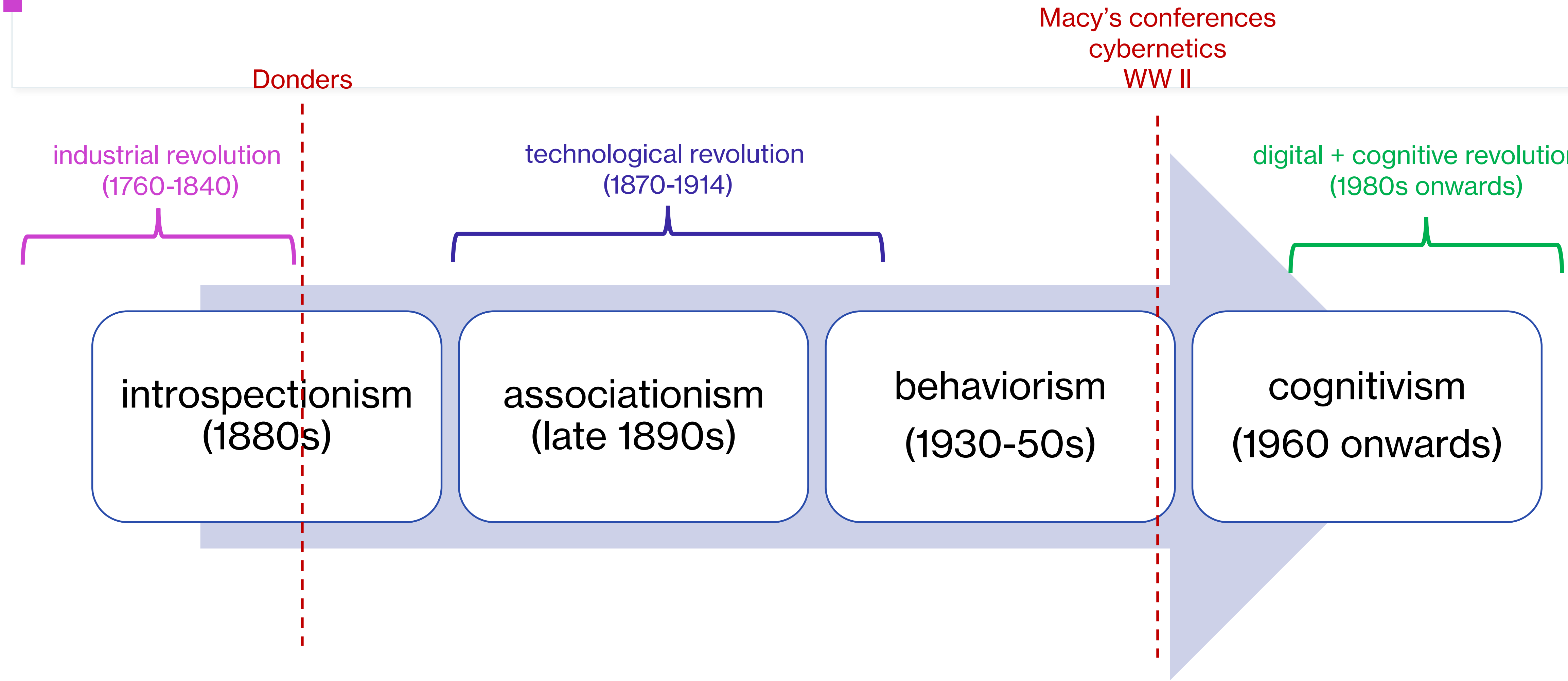
conceptual questions #PRP

I need help clarifying the PRP effect and how it relates to the RW model. The PRP effect is that when 2 identical stims are presented with a short delay between them, the 2nd stim will be harder/take longer to respond to. The RW model says that with practice, our predictions get better, and therefore our learning increases. This would be the opposite of the PRP effect, because instead of allowing for prediction to learn (RW model) the surprise element of second identical stimuli so quickly after does not allow us to apply our learning?

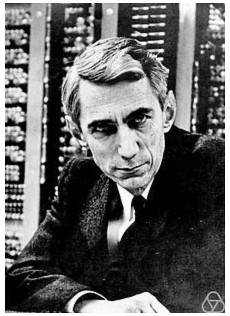
I think that this has to do primarily with the shorter delay implicated in the PRP effect. Something about the occupation of the brain with the first stimulus prevents it from quickly reacting to the second stimulus. Perhaps the short delay doesn't allow for learning the way a longer delay might. I'm curious, how long would the delay have to be for reaction time to get shorter? I'm sure it's dependent on the situation, but have there been any follow-ups to the framework of the PRP effect?

important to separate reaction time (PRP) from learning (RW): the PRP effect is about the slowing down of response time and makes no predictions about whether or not an association is formed between the stimuli and response, whereas the RW model is about the learning of associations between stimulus and responses and makes no predictions about the time taken to learn the association

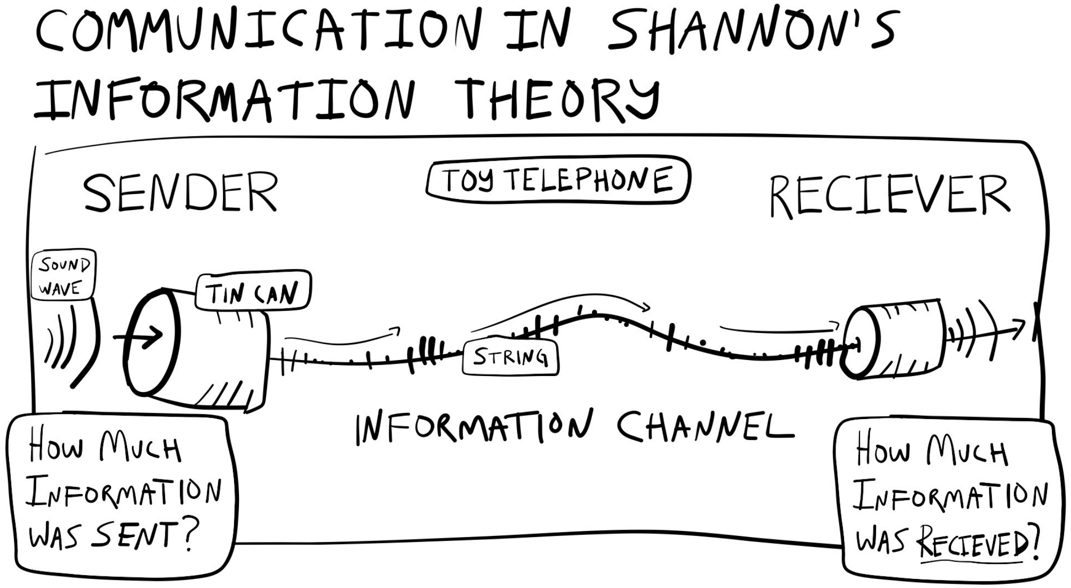
the timeline so far



origins: information theory

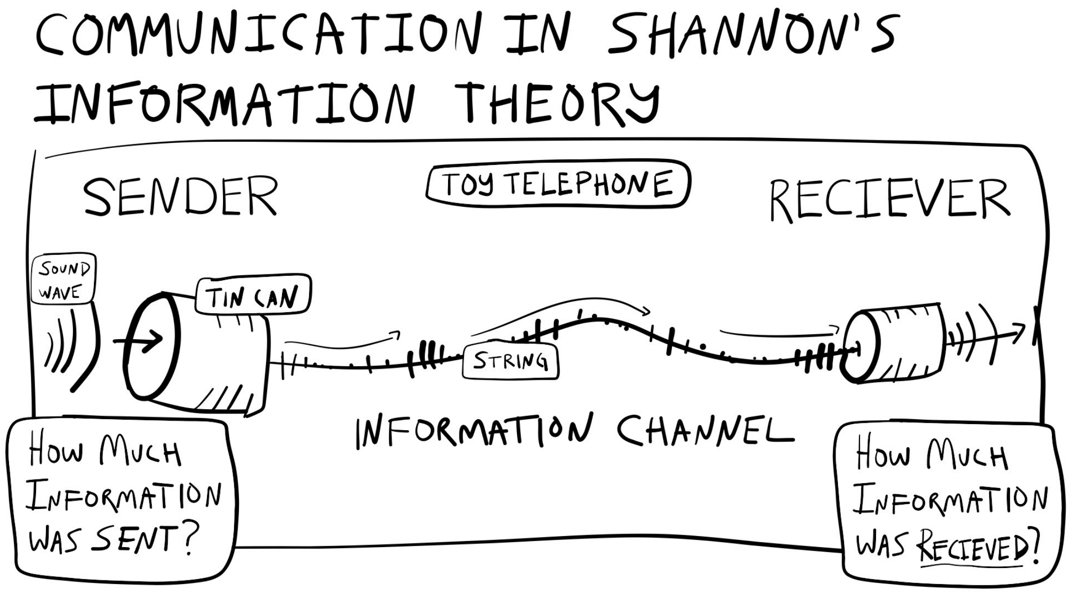


- Claude Shannon, American mathematician regarded the “father of information theory”
  - also contributed to cryptanalysis during WWII
- the main purpose of information theory was to characterize communication systems, not understand cognition/psychology
- but researchers began to see parallels between the idea of an information channel in telecommunications and a channel for transmitting information via mental operations



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
- slightly different intuitions about information...predictability is key here, 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 * sum (0.5(-1) + 0.5(-1)) = -1 * -1
  - H(X) = 1

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

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 * 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

$log_2 P(x_i)$: 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
- 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 × 2) = 4

2^3 = (2 × 2 × 2) = (4 × 2) = 8

conceptual question #bits

- I'm slightly confused on the concept of bits of information. The textbook shows that the relationship between number of bits and number of unique events they can code is defined by raising 2 to the number of bits. However, Crump then gives an example of a sender only being able to send one of four events: A, B, C, or D, saying that 4 bits are are needed to represent these four events. Why is the answer not 2 bits (since 2^(2 bits) = 4 events)?
    - correct!
    - typo in textbook! has been corrected

conceptual question #H

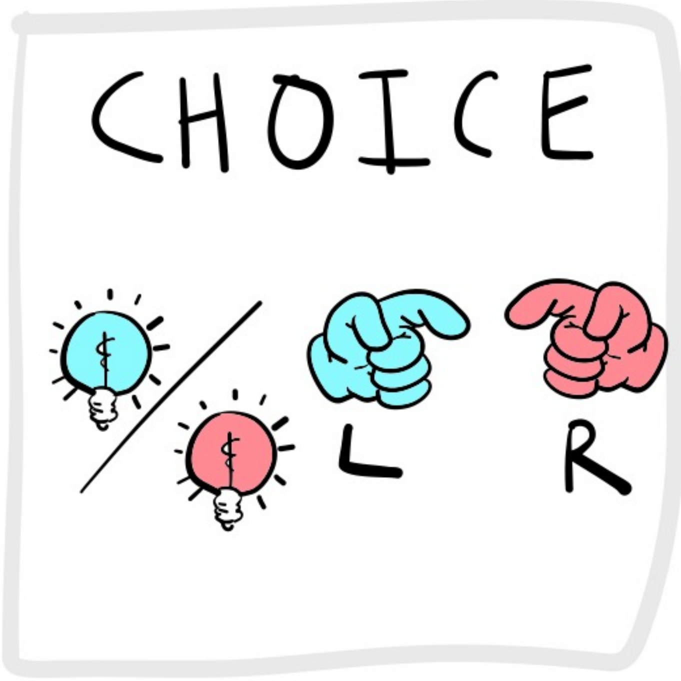
- I am confused about Shannon's H – both the workings of the formula but also the general significance of the equation. What does H mean, what is the general range of values this equation produces and what do different numbers on this scale tell us about the cognitive process/function to do with the information processing?
  - H quantifies the information noise in a system, which is useful when thinking about physical and mental systems: still an actively used concept to quantify uncertainty at different levels, in CS/machine learning, linguistics, cryptology, digital communications and signal processing, etc.

a communication game

- suppose you (sender) had to transmit the outcome of a dice roll to your friend (receiver)
  - you have 6 possible events to communicate
  - 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

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 2

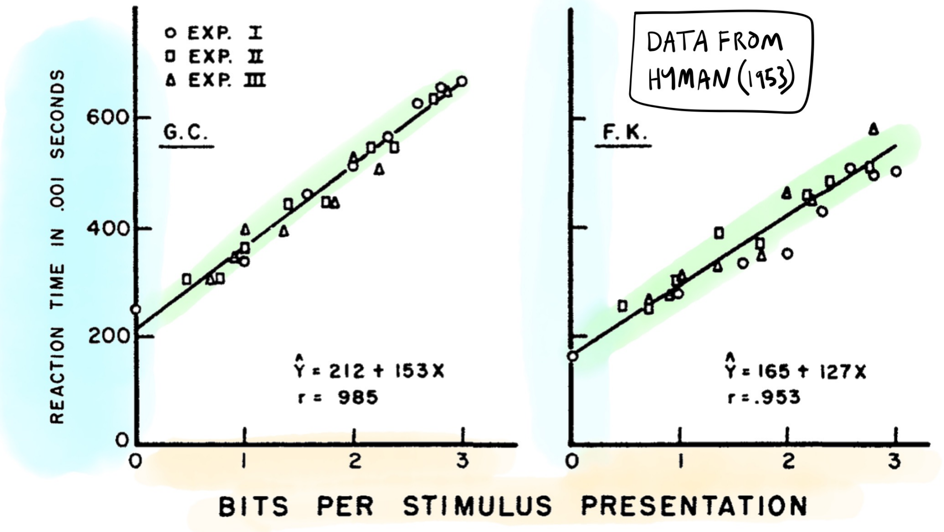
DIFFERENT AMOUNT OF BITS

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)

conceptual question #laws

- As we have discussed a bit in class, "laws" have been mostly phased out of psychology with more nuance and consideration of different environmental, social, and developmental factors. The Hick-Hyman law is an example of one of these older concepts that were viewed as a law but has now seen more nuance and is part of a broad field of study. What has become the new alternative with the acceptance that nothing is truly universal? It seems that certain labs and researchers hold onto hypotheses and attempt to find evidence supporting their claims. Are claims and hypotheses the modern alternative and what value is lost or gained with diminishing universality?

conceptual question #knowing

In this unit about information processing as well as others, there has been a general consensus that not many true laws can be established and that many possible explanations and perspectives exist for a certain topic. If that is the case, how do we know when we truly "know" something in cognition? Are there established facts in the field of information processing now (such as that it occurs in a serial vs. parallel manner)?

Perspectives on Psychological Science
2017, Vol. 12(6) 1100–1122

DOI: 10.1177/1745691617693393
www.psychologicalscience.org/PPS
SAGE

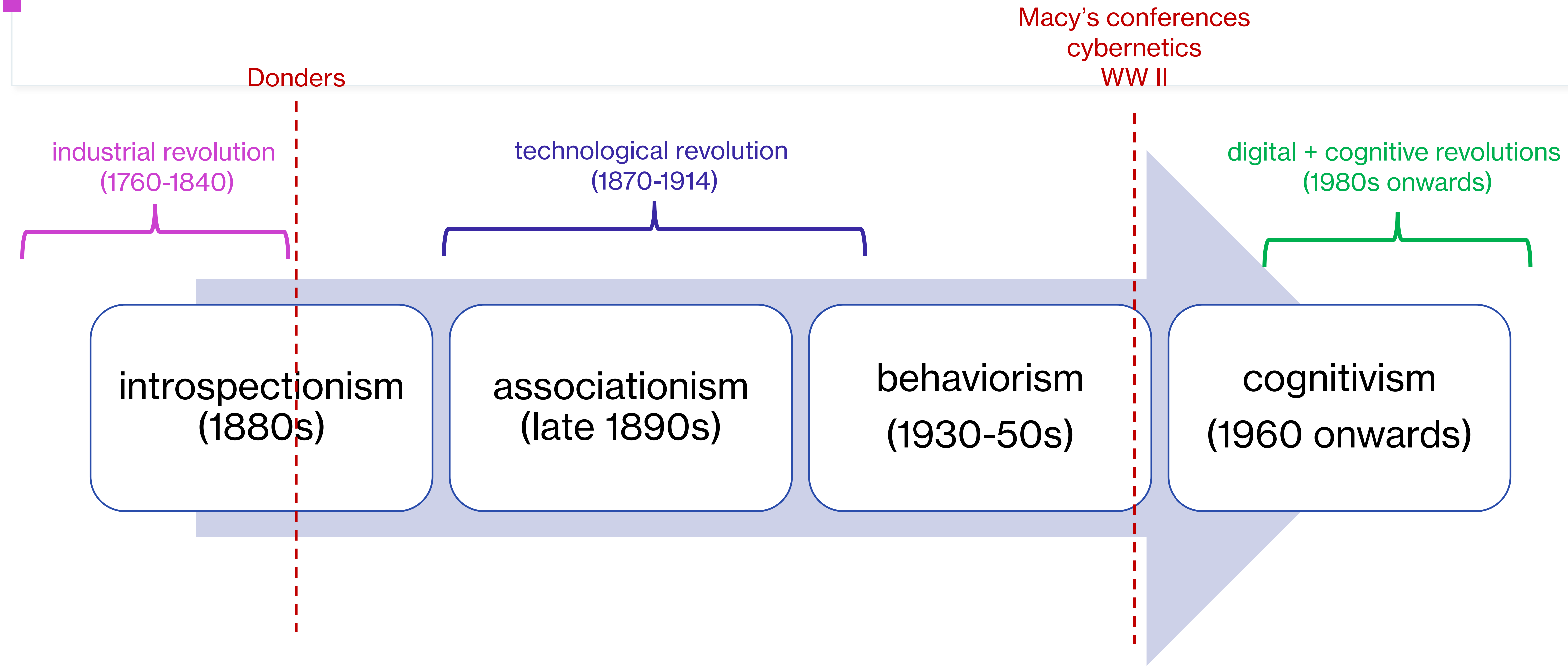
Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning

Tal Yarkoni and Jacob Westfall
University of Texas at Austin

Turing's "thinking" machines

- Alan Turing wrote a paper about whether or not machines can "think" in 1950 where he proposed the imitation game (AKA The Turing Test)
- the broad idea was that if a machine and a human are asked questions by a third interrogator, and the interrogator is unable to decipher which one is human and which is the machine
- the paper has been highly influential in setting the tone for AI as well as the study of cognition
  - ChatGPT is a current real-life example of this idea
  - writing assignment #2 this week explores Turing's paper!
- where do you see Turing's ideas fall in the spectrum if ideologies we have discussed so far?

the timeline so far



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
  - *work on:* project milestone #3
  - *post*: conceptual question
  - *block out time*: practice assessment 1
  - *fill out*: practice assessment survey
- **during** class:
  - structured + unstructured (bring questions) review

logistics: practice assessment + review

- practice assessment 1 is up on Canvas: Modules > Assessment
  - L0 (getting started) – L6 (information processing)
  - 30 multiple-choice + 10 short-answer questions
- answers will be released by the end of this week
- use the weekend to do the assessment in “exam” mode
- use review session + class time to ask questions!
  - Monday and Wednesday, 7-8 pm
- provide us with feedback on the timing/difficulty
  - https://forms.gle/pkcfEvxFC4EXntBX8
  - link also on Canvas > Assessments