Information Theory (following ideas of Claude Shannon)

- We want to think about information as "what we learn"
  - Suppose we are flipping a coin. What do we learn about the probability of heads vs tails each flip if we don't know?
    + For even probability, we learn a moderate amount each flip. Over a long time, we gain confidence in 50% odds.
    + But suppose $P(\text{heads}) \gg P(\text{tails})$. We don't learn much with each head, but we learn a lot from each tail

    Ex We toss ~~H~~ T H H H ... The second time we toss a tail tells us a fair bit about $P(T)$ but each H tells us $P(H) \sim 1$.

  + Another way to think about it is cryptography. It's fast to decode common letters, but rare ones give more information.
  - In data compression, consider a way to make rare letters longer.
    + Suppose we have 4 letters; a, b, c, d.

      We could encode these in binary: a=00, b=01, c=10, d=11.
    + But what if the probabilities are uneven? $P(a)=\frac{1}{2}, P(b)=\frac{1}{4}, P(c)=P(d)=\frac{1}{8}$?
    + Then a=0, b=10, c=110, d=111 uniquely encodes any word. (zero indicates "move to next letter" + no letter has more than 3 digits)
    + What's the average letter length?

      1st code $\langle \text{Length} \rangle = 2$ since all letters are 2 digits

      2nd code $= (\frac{1}{2})(1) + (\frac{1}{4})(2) + (\frac{1}{8})(3) + (\frac{1}{8})(3) = \frac{7}{4}$ smaller!
    + The idea of "letter length" leads to the Shannon information

      $$i(\text{letter}) = -\log_2(P(\text{letter}))$$ in "Shannon bits"
    + The average length is the Shannon entropy

      $$S = \sum_{\text{letters}} -P\log_2(P)$$ Relates to statistical mechanics.

Calculate for a 2-letter system

- Quantum mechanics leads to some extra considerations
  - What if our quantum coin is a spin ½ particle?
    - If we have state $|\uparrow\rangle$ and measure S_z, we get spin up ("heads") 100% of the time, ⟹ there is no information
    - But if you have state $|\uparrow\rangle$ and measure S_x, you get "heads" (positive spin) and "tails" 50% each.
    - Information content should be independent of the question asked.
  - Leads to the distinction of a "pure state" vs. "mixed state"
    - Pure State: a normalized superposition of basis states (what we've been using all year)
      Mixed State: A set of pure states, each of which has a classical probability of being the state of the system
    - Quantum info is basically the Shannon info for each pure state.