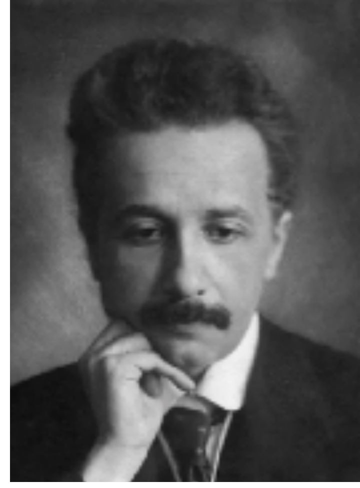
Why use randomness for randomness' sake?



"God does not play dice …"
- Albert Einstein

- Is there a good justification for modeling deterministic processes as random ones? …

- The argument of whether or not some process or phenomenon is truly random is often a philosophical matter. It is a matter of fact, though, that deterministic processes can, to those who only partially observe them, appear random.

- In the world of NLU, this is particularly relevant. Language is a coarse proxy of the human mind, and as such, building a deterministic model that mimics human perception and behavior is unrealistic. The process that generates human language is partially observable (text/speech); but we don't yet have a good model of human intelligence. Building a simple model that is uncertain about its belief is currently our best modeling approach.

- Dirty secret: Our decision to model $p(w \mid context)$ is also one of convenience: we have the internet and labels are free. 

Random variables

- Examples: UV index, temperature, hair color, skin color, coin flip result etc..
- Probability distribution
  - Definition: How likely a random variable (or set of RVs) assumes each of its possible values. We call this a probability density function (pdf) for continuous RVs, and a probability mass function (pmf) for discrete valued RVs. In this class we are primarily concerned with PMFs.
  - A PMF, $P(x)$, must be:
    - Bounded: $\forall x \text{ } in \text{ } X : 0 \leq P(x) \leq 1$
    - Normalized: $\sum_{x \in X} P(x) = 1$