Recall (Lecture XIV)

(Adopt methodology of cognitive science to provide a general framework for philosophy of science [Giere] or to account for the problem of conceptual change therein [Nersessian])

- **Constructive Realism:** "The important relationship between [cognitive] models and the world is not a [logical] semantic relationship, such as truth, but similarity between two nonlinguistic entities, an abstract model and a real system." (Giere, 20)

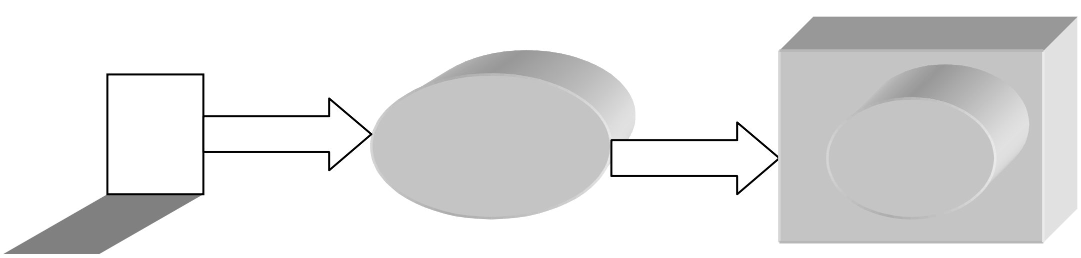
Analogies aren't *arguments*, but no-linguistic *modes of abstraction*. (Nersessian) See her naturalistic classification of the ontologies of cognitive models, underwritten by analogy (a cognitive sub-linguistic activity), in her analysis of Maxwell's diaries (sections 4-6)

"[A] mental model is a **structural analog** of a **real-world or imaginary situation**, **event**, or **process that the mind constructs in reasoning**." (-Nersessian, p. 11)

"What it means for a mental model to be a structural analog is that it **embodies a representation of the spatial and temporal relations among**, and the **causal structures connecting the events**[1] and entities depicted and whatever other information is relevant to the problem-solving task." (ibid.)

"[P]hilosophers consider the processes of conceptual change as mysterious and unanalyzable. Conceptual innovation is held to occur in sudden flashes of insight, with new concepts springing forth from the head of the scientist like Athena, fully grown. This does accord with **retrospective**[2] accounts of **some** scientists, but **if one examines their deeds—their papers, diaries, letters, notebooks—these records support a quite different interpretation in most cases…conceptual change results from extended problem-solving processes…constitut[ing] forms of model-based reasoning: analogical, visual, and simulative modeling."** (13-14)

Analogical model "represents what is common among members of specific classes of physical systems, viewed with respect to a problem context." (16)



Source (known) Target (unsolved problem)

[1] Recall Hacking's notion of experimental intervention (an 'externalized' version of this innatist view.)
[2] Read: "reconstructed"

NOTES ON ELEMENTARY PROBABILITY AND BAYES' THEOREM

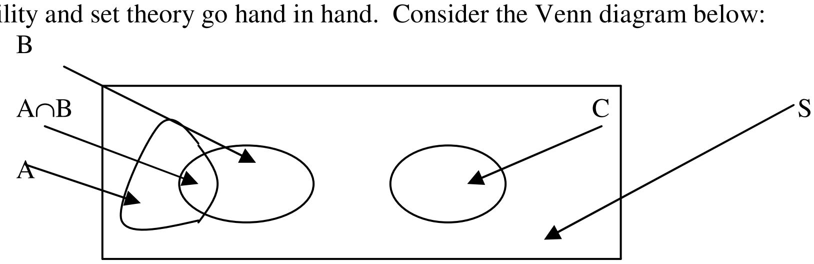
As mentioned[3], ***classical probability*** is the oldest concept, and pertains to those special situations where the sample space (collection of all possible outcomes) **is fixed (*a priori*)** and **each outcome is equally likely to occur.**[4] Its drawback lies in its limitation: such situations usually only occur in controlled and constructed circumstances, (the classical example being games of chance). **For the sake of simplicity, all concepts up toand including Bayes' Thm will be introduced using *classical* probability.**

Empirical probability is most 'statistical,' as one is interpreting a probability as a frequency ratio for any statistical experiment. The sample space is typically a ***histogram***. No 'apriori' assumptions are made about the likelihood of events occurring. Their likelihood is merely their measured frequency in a particular experiment. Therein lies the danger: often attempts are made to extrapolate an empirical probability outside the bounds of the experiment from which it was obtained. This is unwarranted.

- For example, in the case of profiling: Suppose the statistics reveal that in a given neighborhood *A* 16% of all 14-year olds have tried marijuana. Then, I am basing the 16% probability that the fourteen year old from neighborhood *A* has tried pot on an empirical study, and (to a level of confidence/significance) this is a ***justified*** use of empirical probability. However, if I extrapolate that the odds are 16% that *any* fourteen year old has tried pot (regardless where s/he lives) based on ***this*** study, then I'm making a ***conjecture*** (forming a ***subjective*** probability). It's not empirical. This is a general way to distinguish between empirical and subjective: **subjective probability is a measure of belief** and in that cases one may extrapolate outside the given boundaries set by a case study: empirical probability must stay within the study's bounds.

I. ADDITION RULES OF PROBABILITY

Probability and set theory go hand in hand. Consider the Venn diagram below:



- The 'universal set' (rectangle) is the sample space S
- Subsets A,B,C are events in S. A point in S is an outcome.
- Notice that events A and B overlap. That means they're not mutually exclusive. On the other hand, events A,C and events B,C don't overlap. A,C and B,C are examples of mutually exclusive events
- One interprets the probability as a kind of 'area.' (between 0 and 1) It has the following bounds:

  $P(\varnothing) = 0$ $\qquad\qquad\qquad$ $P(S) = 1$

  where: $\varnothing$ is the empty set, or 'impossible event' (an event with no outcomes is a set with no points) and S is the entire sample space. Hence, in general, for any event E: $0 \le P(E) \le 1$

[3] Posted notes on Hawthorne
[4] **Note:** These are quite strong assumptions!

If we want to consider more than one event, we use the connectives: $\cup$, $\cap$ which correspond to 'OR' (always in the inclusive sense) and 'AND.' (Set theory refers to these connectives as 'union,' 'intersection')

The **addition rule** for probability says: For any events A, B in a sample space S:

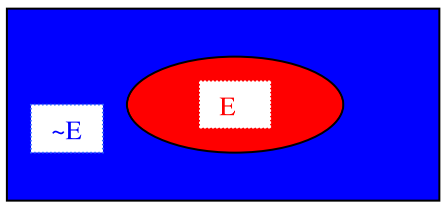
$$P(A\cup B) = P(A) + P(B) - P(A\cap B)$$

which loosely speaking, says: 'area of A union B = their sum (of areas) minus their overlap' (It's necessary to subtract $P(A\cap B)$ from $P(A)$ and $P(B)$ otherwise we're accidentally counting the area twice! (See above Venn Diagram))

- If A and B are mutually exclusive, then: $A\cap B=\varnothing$, so the addition rule says:

$$P(A\cup B) = P(A) + P(B) - P(\varnothing) = P(A) + P(B) \qquad \text{(Since } P(\varnothing) = 0\text{)}$$

Certainly for any event E: E and $\sim E = \{x \mid x \notin E\}$ are **mutually exclusive.** Furthermore: $E\cup\sim E=S$. See Venn Diagram below:



The oval region represents some event E. The blue region is ~E. Certainly the union of E with ~E is the entire sample space S, i.e.: $E\cup\sim E= S$. However, E and ~E share no region in common, (they are mutually exclusive) so: $E\cap\sim E=\varnothing$

Hence using the addition rule:

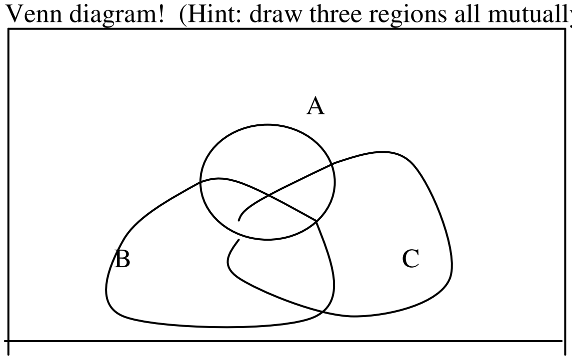
$$P(S) = P(E) + P(\sim E) - 0 = P(E) + P(\sim E)$$

therefore: $P(\sim E) = 1 - P(E)$

For more than two events the addition rule becomes complicated. Try verifying:

$$P(A\cup B\cup C)=P(A)+P(B)+P(C) -P(A\cap B) -P(B\cap C) - P(A\cap C) + P(A\cap B\cap C)$$

using a Venn diagram! (Hint: draw three regions all mutually overlapping)



1. Observe that when adding the 'areas' (probabilities) of regions A,B,C we accidentally add the area of of the regions of overlap between A and B (denoted $A\cap B$), between A and C (denoted $A\cap C$), between B and C (denoted $B\cap C$) **twice.**
2. Observe that when doing this we accidentally add the area that all three regions share in common (denoted $A\cap B\cap C$) **thrice** (three times.)
3. Therefore, to compensate for the mistake we made in 1., we must subtract that extra area we added between the respective regions A & B, A & C, B & C to get: $P(A)+P(B)+P(C) - P(A\cap B) - P(B\cap C) - P(A\cap C)$.
4. Unfortunately, however, when doing that (in 3.) we took away the area of overlap of the three region ($A\cap B\cap C$) **three times**, which leaves us with no area anymore in this region (as we've seen in 2. that area was only added three times.) Therefore we need to add $P(A\cap B\cap C)$ in the above expression in 3. to get our final answer: $P(A\cup B\cup C)=P(A)+P(B)+P(C) - P(A\cap B) - P(B\cap C) - P(A\cap C) + P(A\cap B\cap C)$

Example: I have a deck of 52 cards. I draw one card. Calculate the probability that the card will be of a black suit or a clubs or a 10

Let event A = {black suit} B={club} C={jack}

Then:

$$P(A\cup B\cup C)=P(A)+P(B)+P(C) - P(A\cap B) - P(B\cap C) - P(A\cap C) + P(A\cap B\cap C)$$
$$= {}^{26}/_{52} + {}^{13}/_{52} + {}^{4}/_{52} - {}^{13}/_{52} - {}^{1}/_{52} - {}^{2}/_{52} + {}^{1}/_{52} = {}^{28}/_{52} = {}^{7}/_{13}$$

(Observe some of the repetition of the probabilities in the above example, this gives us an insight into the notion of **independent events.)**

II. MULTIPLICATION RULE

So far we've only looked at events taking place **at one instant.** Though interesting results occur (described by addition rule) we'd also like to explore what happens when we have **two or more events taking place at different times.** Think back to chapter 4. We drew trees illustrating outcomes. Reading the tree 'across' (horizontally) made us think of outcomes separated by 'OR.' On the other hand reading the tree 'down' made us think of generations, separated by 'AND.' (More precisely: 'AND THEN.') We devised a multiplication rule for 'AND THEN' and addition rule for 'OR.' We just saw what happens to the addition rule when we think of probability. We're about to explore what happens when we bring in probability when considering multiplication.

Definition:

- Suppose event A occurs at instant T1. Suppose event B occurs at instant T2 (where T2 > T1, i.e., event B follows event A.) We say events A and B are **independent** if the probability of event A happening won't affect the probability of event B happening. Otherwise, we say they're **dependent**.

So, we see we need at least two events occurring at different instants to understand dependence/independence. On the other hand, mutual exclusivity has to do with events sharing or not sharing outcomes **at the same instant.** It's important not to confuse the two ideas. The table below lists some examples:

(EXAMPLES)	**2 INDEPENDENT EVENTS**	**2 DEPENDENT EVENTS**
2 MUTUALLY EXCLUSIVE EVENTS	Drawing a King and then drawing a Jack in a deck of cards, where the sampling is **with replacement**.	Drawing a King and then drawing a Jack in a deck of cards, where the sampling is **without replacement**.
2 EVENTS NOT MUTUALLY EXCLUSIVE	A couple has two children over a two-year period. Let: Event 1 = {boy, girl}. Let Event 2 = {girl, girl}	A cook prepares a meal. Let: Event 1 = {boil the water} Event 2 = {Steam the vegetables}

The Multiplication Rule for Probability (for two events) is stated as

a.) If events E1, E2 are independent, then: $P(E1 \cap E2) = P(E1)P(E2)$

b.) If events E1, E2 are **not** independent, then: $P(E1 \cap E2) = P(E1)P(E2|E1)$

- $P(E2|E1)$ is called **conditional probability**. It means: 'the probability of E2 ***given that*** E1 has occurred.'
- Interestingly, the notion of dependence doesn't necessarily need to follow a strict, sequential time order, since it's based on logical AND ($\wedge$)[5]. **Bayes' Theorem**, which has a lot to do with ***counterfactual*** reasoning[6], exploits this **commutative property of** $\wedge$. So it's also true that: $P(E1 \cap E2) = P(E2 \cap E1) = P(E2)P(E1|E2)$
- The above rule gives us a convenient formula for calculating conditional probability:

$P(E2|E1) = {}^{P(E1 \cap E2)}/_{P(E1)}$ $\qquad$ $P(E1|E2) = {}^{P(E1 \cap E2)}/_{P(E2)}$

(Another way of looking at this is: $P(E1 \cap E2) = P(E2 \cap E1)$ ('AND' need not follow strict sequential time-order.)

***Example*: Consider the midterm exam. Let E1 = {A letter grade}, E2 = {Took version 3}**

Then: $P(E1|E2) = {}^{P(E1 \cap E2)}/_{P(E2)} = {}^{(2/24)}/_{(8/24)} = {}^{1}/_{4}$

This answer can obviously directly established, since out of the 8 who took version 3, 2 got an A, i.e. $P(E1|E2) = {}^{2}/_{8} = {}^{1}/_{4}$

The Multiplication Rule for Probability (for three events) is stated as

a.) If events E1, E2, E3 are independent, then: $P(E1 \cap E2) = P(E1)P(E2)P(E3)$

b.) If events E1, E2, E3 are **not** independent, then: $P(E1 \cap E2 \cap E3) = P(E1)P(E2|E1)P(E3|E1 \cap E2)$

Example: Suppose I draw two cards **without replacement.** Find the probability of getting a club **at least once.**

---

[5] Recall **Lecture II.**

[6] An example of a counterfactual (which seems to 'go back in time,' in the sense that the flow of the argument is from conclusion to premise.)

I) I didn't wait for you (FACT)
II) Because I didn't know you were coming. (INFERENCE)
Therefore: III) If I knew you were coming, I would have waited. (COUNTERFACTUAL)

Let E1 = {Club on first draw)   E2 = {Club on Second draw}

- E1 and E2 are neither independent nor mutually exclusive. To calculate the probability of getting a club at least once, we're asking:

  P(E1 or E2) = P(E1) + P(E2) - P(E1 and E2)

- But according to the Multiplication Rule, since E1 and E2 are dependent: P(E1 and E2) = P(E1)P(E2|E1)

So: P(E1 or E2) = P(E1) + P(E2) - P(E1)P(E2|E1)

- ***We have to very careful with P(E2)!*** (Because E1 and E2 are not independent.) Consider, for example **if** we had drawn a club the first time (E1). Then $P(E2) = {}^{12}/_{51}$. On the other hand, if we had **not** drawn a club first (~E1), then: $P(E2) = {}^{13}/_{51}$. So how do we resolve this ambiguity? By the Addition Rule.

  P(E2) = P(E1 and E2) + P(~E1 and E2)

  (Note that E1 and not-E1 are obviously mutually exlusive!)

  And once again, because E1 and E2 are dependent, using the Multiplication Rule:

  P(E2) = P(E1)P(E2|E1)+ P(~E1)P(E2|~E1)

Therefore substituting :

P(E1 or E2) = P(E1) + P(E1)P(E2|E1)+ P(~E1)P(E2|~E1)- P(E1)P(E2|E1)
= P(E1) + P(~E1)P(E2|~E1)
$= {}^{13}/_{52} + ({}^{39}/_{52})({}^{13}/_{51}) = {}^{(13*51 + 39*13)}/_{52*51} \approx 0.44$ (44%)

III.BAYES THEOREM

Suppose a rare diseases infects 1 in 10,000. Suppose there's a test which is 99% accurate. Suppose one tests positive. What are the chances of being sick?

Let: P(S) = Probability of infection = 0.0001 then
P(~S) =Probability of not being infected =0.9999

then P(+| S) = Probability of testing positive given that one is infected = 0.99
P(+| ~S) = Probability of testing positive given that one is **not** infected = 0.01 ('false positive')

Given above, the question becomes:

Find P(S| +) (Find probability of being infected given one has tested positive for the disease)

From multiplication rule: $P(S|+) = \dfrac{P(S \cap +)}{P(+)}$

Also, from the addition rule: $P(+) = P(S\cap+) + P(\sim S\cap+)$

(the event of testing positive can be split into two disjoint events: testing positive and being infected OR testing positive and not being infected)

…And from the multiplication rule: $P(S \cap +) = P(+|S)P(S)$ $P(\sim S \cap +) = P(+|\sim S)P(\sim S)$

Therefore, adopting the above substitutions, we can write:

$$P(S|+) = \frac{P(+|S)P(S)}{P(+|S)P(S) + P(+|\sim S)P(\sim S)}$$

This formula is called **Bayes' Formula**. Though complicated, observe that we have expressed our question entirely in terms of the information given. In other words, we've answered our question! The answer is:

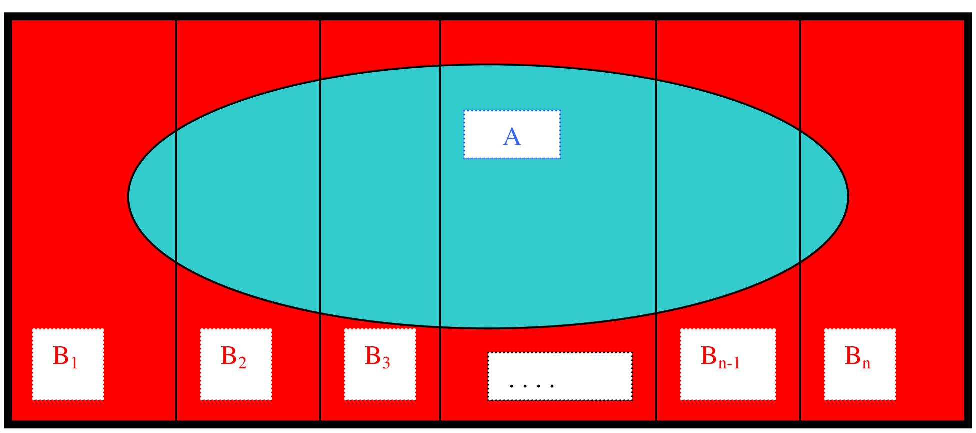
$$P(S|+) = \frac{(.99)(.0001)}{(.99)(.0001) + (.01)(.9999)} = \mathbf{0.009803}$$

In other words, if a rare disease strikes one in 10,000 and you test positive for it, and the test is 99% accurate, this means your chances of actually being infected are 0.0098 or less that 1%! This is just one of the many counterintuitive answers that arise in Bayesian probability.

Actually, when you think carefully, the answer is **not** that surprising. Why? because a test that is 99% accurate means it has a 1% margin for error. Therefore if you test positive, all you can conclude is that your chances of actually being infected have increased by a factor of 100. If the disease strikes one in 10,000 that means your chances of being sick have increased by 100× , i.e. the chances of being sick in the case of testing positive are now one in 100.

Bayes' Formula:
For (disjoint) events $B_1, B_2, B_3, \ldots, B_n$, where: $B_1 \cup B_2 \cup B_3 \cup \ldots \cup B_n = S$

$$P(B_k \mid A) = \frac{P(B_k)P(A|B_k)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \quad \ldots \quad + P(B_n)P(A|B_n)} \qquad \text{for } 1 \le k \le n$$



Example: In a factory, machines I,II,III produce springs of the same length. Of their production, I,II,III produce 2,1,3% defective springs. Machine I produces 35%, II produces 25%, III produces 40%.
If one spring is selected from the total produced in one day, and found to be defective find probablility that it was produced by III

Let D mean 'defective,' ~D mean 'not defective'

Then: $P(D|I) = 0.02$ $P(D|II)=0.01$ $P(D|III) = 0.03$

$P(I) = 0.35$ $P(II) = 0.25$ $P(III) = 0.40$

Find: $P(III|D)$ (the probability part was produced by III given that the part is defective)

According to Bayes' Formula:

$$: P(III|D) = \frac{P(D|III)P(III)}{P(D|I)P(I) + P(D|II)P(II) + P(D|III)P(III)}$$

$$= \frac{(.03)(.4)}{(.02)(.35) + (.01)(.25) + (.03)(.4)} = 0.558 \text{ or } \sim 56\%$$

Exercises[7]**:**

1. Given the posted midterm scores, let V1={took version 1}, V2= {took version 2}, and V3 = {took version 3}. Obviously V1, V2, V3 are mutually exclusive.
   Let B ={score| 80% ≤ score < 89%}. A.) (1 pt) Find ~~$P(B | V2)$~~ $P(V2|B)$ directly
   b.) (3 pts) Use the Bayes formula to verify your answer.

2. (4) Suppose Box I contains 3 red and 4 blue marbles and Box II 2 red and 7 blue marbles. Suppose someone tosses a coin. If it's heads, a marble is chosen from Box I. If tails, a marble is chosen from Box II. Suppose whoever tosses the coin doesn't reveal whether it turned up heads or tails (assume it's a fair coin), but s/he does reveal that a red marble was chosen. Find the probability (using Bayes' formula) that Box II was chosen.

3. (2) A box contains 6 red and 5 white marbles. Two are drawn successively from the box without replacement, and it is noted that the second one is red. What is the probability that the first one is also red? (You don't need to use Bayes' formula for this problem)

[7] If you hand these exercises in on Thursday, October 25, you can up to 10 points extra credit on your midterm score. For maximum credit, however, you must show work in legible detail, in a manner similar to the detail presented in these notes.

<u>Bayesian Epistemology:</u> Extending the strategy to encompass *open-ended* situation (where sample space isn't closed): Central problem: Justifying reliability of switching conditionals (posterior/prior) probabilities.

<u>Bayesian Confirmation Theory</u>: Evidence *E* confirms hypothesis *H* (to a degree of inductive support[8]) if prior probability $P(H|E) > P(H)$

Popper's critique: Assumption of 'logical omniscence' (Assumption that degrees of belief satisfy probability laws)

Sources:

http://plato.stanford.edu/entries/epistemology-bayesian/

http://plato.stanford.edu/entries/bayes-theorem/

http://www.nyu.edu/gsas/dept/philo/user/strevens/Classes/Conf06/BCT.pdf

[8] A notion, for those who wrote paper #1, which is not free from controversy!