Philosophy 148 — Announcements & Such

- Overall, people did very well on the mid-term ($\mu = 90, \sigma = 16$).

- HW #2 is still being graded. We’ll have those ready Thursday.

- Recall: we’re using a straight grading scale for this course. Here it is:
  - A+ > 97, A (94,97], A- (90,94], B+ (87,90], B (84,87], B- (80,84],
    C+ (77,80], C (74,77], C- (70,74], D [50,70], F < 50.

- I’ve posted solutions for HW #1 (HW #2 solutions are coming soon).

- There is a HW #3 discussion tonight @ 6pm @ 251 Dwinelle.
  HW #3 is due Thursday. HW #4 will be assigned on Thursday.

- Today: a brief Review of Joyce’s argument. Then, “Logical” Probability:
  - Leibnizian/Wittgensteinian Roots of “Logical” Probability
  - The Principle of Indifference
  - Carnap’s (Early) Formal Theories of “Logical” Probability
Review of Joyce’s Epistemic Argument for Probabilism

- Joyce aims to show that, unless $q$ is a probability function, there will exist a $q'$ which is closer to the truth - no matter what the truth turns out to be.

- The “closeness to the truth” of $q$ in world $w$ is based on some measure $\rho$ of the distance between $q(X)$ and the truth-value ($w(X) = 0$ or $1$) of $X$ in $w$, summed over all of the $X$’s in $S$’s set of entertainable propositions $B$.

- Joyce’s argument rests on four assumptions $\mathcal{A}$ about the inaccuracy measure $I$. $\mathcal{A}$ is consistent with $\rho^\dagger(q(X), w(X)) = |q(X) - w(X)|^2$, but not $\rho^*(q(X), w(X)) = |q(X) - w(X)|$, or various other distance measures.

- Maher shows that $\rho^*$ violates two of Joyce’s axioms: (WC) and (S), and he argues that $\rho^*$ is nonetheless a reasonable measure of “distance from truth”. Maher also shows that Joyce’s Theorem fails if one uses $\rho^*$.

- Joyce’s defense of $\rho^\dagger$ appeals to “Expected Distance from Truth” of $q$. I mentioned two problems with Joyce’s reply: (1) it appeals to something that seems pragmatic, and (2) it presupposes that $q(\neg p) = 1 - q(p)$. 

UCB Philosophy  Logical Probability  04/01/08
The motivation for “logical interpretations” of probability has its roots in *inductive logic*. Inductive Logic (more later) is meant to be the science of *argument strength* — a quantitative generalization of entailment (\(\models\)).

Inductive Logic aims to explicate a quantitative measure \(c(C, P)\) of the “degree to which the premises of an argument \(P\) jointly confirm its conclusion \(C\).” Keynes, Carnap, and others worked on such theories.

As we will study more in our confirmation unit, one of the key desiderata of IL is that \(c\) should quantitatively generalize deductive entailment:

\[(D_1)\] The relations of deductive entailment and deductive refutation should be captured as limiting (extreme) values of \(c\) with cases of ‘partial entailment’ and ‘partial refutation’ lying somewhere on a \(c\)-continuum (or range) between these extreme values of the confirmation function.

Exercises: (i) *the conditional probability* \([\Pr(C \mid P)]\) satisfies \(D_1\), but *the probability of the corresponding conditional* \([\Pr(P \rightarrow C)]\) does not.
Another key historical desideratum for inductive logic is:

(D$_2$) Inductive logic (i.e., the non-deductive relations characterized by inductive logic) should be objective and logical.

Carnap on this desideratum:

“Deductive logic may be regarded as the theory of the relation of logical consequence, and inductive logic as the theory of another concept [c] which is likewise objective and logical, viz., . . . degree of confirmation.”

Carnap on the terms “logical” and “objective” as they apply to c.

“The principal common characteristic of the statements in both fields is their independence of the contingency of facts. This characteristic justifies the application of the common term ‘logic’ to both fields.”

“That c is objective means this: if a certain c value holds for a hypothesis (H) with respect to evidence (E), then the value c(H, E) is independent of what any person may happen to think about these sentences.”
Whither Logical Probability? 3

- While $\Pr(C \mid P)$ clearly satisfies $D_1$, it is unclear whether $\Pr(C \mid P)$ satisfies $D_2$. If $\psi(C, P) \overset{\text{def}}{=} \Pr(C \mid P)$, then, we seem to need $\Pr(C \mid P)$ itself to be logical if $\psi(C, P)$ is to be logical. Hence, the need for “logical probability”!

- Moreover, if we assume the standard definition of $\Pr(C \mid P)$, then we need an unconditional probability function that is itself logical. The basic “Leibnizian” idea, which underlies this “interpretation” of probability:

$$\Pr(C \mid P) = \frac{\Pr(P \& C)}{\Pr(P)} = \frac{\text{The proportion of logically possible worlds in which } P \& C \text{ is true}}{\text{The proportion of logically possible worlds in which } P \text{ is true}}$$

- Such unconditional logical probability functions are called logical measure functions ($m$). Intuitively, these are intended to measure the “proportion of logically possible worlds in which a proposition is true”.

- Wittgenstein, Carnap, and others give precise explications of this vague concept of “logical probability”. They work within logical languages $\mathcal{L}$, and they work with descriptions of possible worlds — sentences in $\mathcal{L}$. 
Wittgensteinian Propositional Logical Probability 1

- In the *Tractatus*, Wittgenstein presents the idea of a truth-table for a propositional language $\mathcal{L}_P$. He proposes a logical measure function $m$ on $\mathcal{L}_P$, which assigns *equal probability* to each *state description* of $\mathcal{L}_P$.

- Let $\mathcal{L}_P^n$ be a propositional language with $n$ atomic sentences. The measure function $m$ would assign $m(s_i) = \frac{1}{2^n}$, for all state descriptions $s_i$ of $\mathcal{L}_P^n$.

Here’s what $m$ looks like over an $\mathcal{L}_P^3$ language (this should look familiar!):

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>State Descriptions</th>
<th>$m(s_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>$s_1 = A &amp; B &amp; C$</td>
<td>$m(s_1) = 1/8$</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>$s_2 = A &amp; B &amp; \sim C$</td>
<td>$m(s_2) = 1/8$</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>$s_3 = A &amp; \sim B &amp; C$</td>
<td>$m(s_3) = 1/8$</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>$s_4 = A &amp; \sim B &amp; \sim C$</td>
<td>$m(s_4) = 1/8$</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>$s_5 = \sim A &amp; B &amp; C$</td>
<td>$m(s_5) = 1/8$</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>$s_6 = \sim A &amp; B &amp; \sim C$</td>
<td>$m(s_6) = 1/8$</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>$s_7 = \sim A &amp; \sim B &amp; C$</td>
<td>$m(s_7) = 1/8$</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>$s_8 = \sim A &amp; \sim B &amp; \sim C$</td>
<td>$m(s_8) = 1/8$</td>
</tr>
</tbody>
</table>
Wittgensteinian Propositional Logical Probability 2

• Here are some important technical facts about Wittgensteinian $\mathcal{L}_P^n$-Pr:
  - Every atomic sentence in every $\mathcal{L}_P^n$ probability model has probability $\frac{1}{2}$.
  - Every collection of atomic sentences $\{A_1, \ldots, A_n\}$ in every $\mathcal{L}_P^n$ probability model is mutually probabilistically independent.
  - For every pair of sentences $p$ and $q$ (atomic or compound) in every $\mathcal{L}_P^n$ probability model, $p$ and $q$ are probabilistically dependent iff $p$ and $q$ are logically dependent (i.e., iff $p \models q$ or $\neg p \models q$ or $q \models p$ or $\neg q \models p$).
    [This depends on the equiprobability of the state descriptions. Why?]
  - For all $p$ and $q$ in any $\mathcal{L}_P^n$ model, $\Pr(q \mid p) = 1$ if and only if $p \models q$.
    [This does not depend on the equiprobability of the S.D.’s. Why?]

• The key philosophical question about this kind of approach is:
  - What makes the assignment of equal probability to the state descriptions the “logical” probability assignment? What logical principle is at work here? There are 2 kinds of answers: (i) the Principle of Indifference (PI), and (ii) permutation invariance. Carnap on (PI)…
Carnap on the Principle of Indifference 1

- The Principle of Indifference (PI) says, roughly, that if an epistemically rational agent’s total evidence $K$ does not favor any member of a partition of possible states $\{s_1, \ldots, s_n\}$ over any other member, then that agent’s degrees of credence (assumed to be probabilities!) should satisfy:

\[
\Pr(s_i | K) = \Pr(s_j | K), \text{ for all } i, j.
\]

- The invocation of (PI) in this context should be puzzling for a few reasons:
  - First, (PI) sounds like an epistemic principle about what an agent’s epistemic probabilities should be, under certain circumstances.
  - Should logical principles be determined by epistemic constraints? This is controversial, even for deductive logic (think inconsistent beliefs): *You shouldn’t believe everything if your set of beliefs happens to contain a contradiction (especially, in light of the Preface!).*  
  - So, even for deductive logic, epistemic principles seem to make for odd logical constraints. Why should inductive logic be any different?
• Presumably, the (PI) approach goes as follows. Assume that “all we know is logic”. Call this background knowledge $K_T$. Assume that $K_T$ does not favor any state description over any other. Then, $m$ can be thought of as $\Pr(\bullet \mid K_T)$, and we have $\Pr(s_i \mid K_T) = \Pr(s_j \mid K_T)$, for all $i, j$, as desired.

• Carnap cleverly argues that (PI) is a logical (not epistemic!) principle:

  ... the statement of equiprobability to which the (PI) leads is, like all other statements of inductive probability, not a factual but a logical statement. If the knowledge of the observer does not favor any of the possible events, then with respect to this knowledge as evidence they are equiprobable.

• Here’s how the argument is supposed to go (as far as I can see):

  - Intuitively, “$K$ does not favor $s_i$ over $s_j$ (any $i \neq j$)” $\Rightarrow$ “$K$ confirms $s_i$ to the same degree as $K$ confirms $s_j$ (each $i \neq j$)”, i.e., $c(s_i, K) = c(s_j, K)$.
  
  - If we assume that $c(x, y) = \Pr(x \mid y)$, then logic alone implies that:
    
    $$c(s_i, K) = c(s_j, K) \Rightarrow \Pr(s_i \mid K) = \Pr(s_j \mid K), \text{ for all } i \text{ and } j.$$

  - $\therefore$ Logic alone implies that if $K$ does not favor $s_i$ over $s_j$ (for any $i \neq j$ in a partition of states), then $\Pr(s_i \mid K) = \Pr(s_j \mid K)$, for each $i, j$. □
Carnap on the Principle of Indifference 2

- Carnap’s argument is clever, but ultimately not terribly compelling.
- As we will see next, Carnap’s own Predicate-Logical approach to “logical probability” deviates from the application of (PI) to the state descriptions of his logical languages [and, for reasons that look epistemic, yet again!].
- Aside from these worries about the “logicality” of (PI) and its success at generating the “right ‘logical’ probability models”, there are other problems with Carnap’s argument. Here’s a cursory sketch (more later):
  - Carnap assumes that the correct (probabilistic) explication of the confirmation function is just the conditional probability function, *i.e.*, 
    \[ c(x, y) = \Pr(x \mid y) \]
  - If it were obvious that \( \Pr(x \mid y) \) is the best (probabilistic) explication of \( c(x, y) \), then Carnap’s argument might be sound (if not informative!).
  - We’ll soon see that \( \Pr(x \mid y) \) may not be the best explication of \( c(x, y) \).
- We’ll discuss other aspects of the (PI) in the next units of the course.
Carnapian Monadic Predicate Logical Probability 1

- Generalizing on Wittgenstein, Carnap defined his logical measure functions over sentences in monadic predicate logical languages $\mathcal{L}_{Q}^{m,n}$ containing $n$ monadic predicates ($F, G, \ldots$) and $m$ constants ($a, b, \ldots$).
- To fix ideas, consider the language $\mathcal{L}_{Q}^{2,2}$, which contains two monadic predicates $F$ and $G$ and two individual constants $a$ and $b$.
- In $\mathcal{L}_{Q}^{2,2}$, we can describe 16 states, using the 16 state descriptions of $\mathcal{L}_{Q}^{2,2}$:

  - $Fa \& Ga \& Fb \& Gb$
  - $Fa \& Ga \& \sim Fb \& \sim Gb$
  - $Fa \& \sim Ga \& \sim Fb \& Gb$
  - $\sim Fa \& Ga \& \sim Fb \& Gb$
  - $\sim Fa \& \sim Ga \& Fb \& Gb$
  - $\sim Fa \& \sim Ga \& \sim Fb \& \sim Gb$
  - $Fa \& Ga \& \sim Fb \& \sim Gb$
  - $Fa \& \sim Ga \& \sim Fb \& \sim Gb$
  - $Fa \& Ga \& Fb \& Gb$
  - $Fa \& \sim Ga \& Fb \& Gb$
  - $Fa \& Ga \& \sim Fb \& Gb$
  - $Fa \& \sim Ga \& \sim Fb \& Gb$
Carnapian Monadic Predicate Logical Probability 2

- Following (PI), Carnap’s first measure function $m^\dagger$ assigns *equal probability* to each state description $s_i$ of $\mathcal{L}_Q^{m,n}$. In our $\mathcal{L}_Q^{2,2}$, $m^\dagger(s_i) = \frac{1}{16}$.

- We extend $m^\dagger$ to all $p \in \mathcal{L}_Q^{m,n}$ in the standard way (the Pr of a disjunction of mutually exclusive sentences is the sum of the Pr’s of its disjuncts).

- Since every $p \in \mathcal{L}_Q^{m,n}$ is equivalent to some disjunction of state descriptions, and every pair of state descriptions is mutually exclusive, this gives us a complete unconditional Pr-function $\Pr^\dagger(\cdot)$ over $\mathcal{L}_Q^{m,n}$.

- Finally, we define the conditional probability function $\Pr^\dagger(q \mid p)$ over pairs of sentences in $\mathcal{L}_Q^{m,n}$ (in the standard way) as the following *ratio*: $\frac{\Pr^\dagger(p \& q)}{\Pr^\dagger(p)}$.

- Claims of the form $^\dagger\Pr(q \mid p) = x^\dagger$ are *analytic in $\mathcal{L}_Q^{m,n}$* since their truth-values are determined solely by the syntactical structure of $\mathcal{L}_Q^{m,n}$.

- But, why is *this* choice of measure function $m^\dagger$ *logical*? Logicality is ensured by the application of (PI) to state descriptions. Or is it?
Carnapian Monadic Predicate Logical Probability 3

- As it turns out, Carnap ultimately \textit{rejects} the measure function \( m^\dagger \) in favor of an alternative measure function \( m^* \), for epistemic-sounding reasons.

- Carnap notes that \( m^\dagger \) causes \( \Pr^\dagger \) to have the following property \((b \neq a)\):

\[
(*) \quad \Pr^\dagger (Fb \mid Fa) = \frac{\Pr^\dagger (Fb \& Fa)}{\Pr^\dagger (Fa)} = \frac{4 \cdot \frac{1}{16}}{8 \cdot \frac{1}{16}} = \frac{1}{2} = 8 \cdot \frac{1}{16} = \Pr^\dagger (Fb)
\]

- In other words, \((*)\) says that one object \( a \)'s having property \( F \) can never raise the probability that another object \( b \) also has \( F \). And, this \textit{generalizes} to any number of \( F \)s: \( \Pr^\dagger (Fa_1 \mid Fa_2 \& \cdots \& Fa_m) = \Pr^\dagger (Fa_1) \).

- Carnap (\textit{et al}) characterize this (in \textit{epistemic} terms) as \( m^\dagger \) leading to a function \( \Pr^\dagger \) that fails to allow for \textit{“learning from experience”}. This is \textit{epistemic and diachronic}, since it assumes \textit{learning-by-conditionalization}.

- Carnap views this consequence of applying (PI) to the state descriptions of \( L_{Q}^{m,n} \) as unacceptable. Does he then think that \( K_T \) \textit{“favors”} some state descriptions over others. But, which ones? Enter the \( m^* \) measure \ldots
Carnapian Monadic Predicate Logical Probability 4

- Two state descriptions $s_i$ and $s_j$ in $L_Q^{m,n}$ are permutations of each other if one can be obtained from the other by a permutation of constants.

- "Fa & ~Ga & ~Fb & Gb" can be obtained from "~Fa & Ga & Fb & ~Gb" by permuting "a" and "b". Thus, "Fa & ~Ga & ~Fb & Gb" and "~Fa & Ga & Fb & ~Gb" are permutations of each other (in $L_Q^{2,2}$).

- A structure description is a disjunction of state descriptions, each of which is a permutation of the others. $L_Q^{2,2}$ has 10 structure descriptions:

  - $(Fa & ~Ga & ~Fb & Gb) \lor (~Fa & Ga & Fb & ~Gb)$
  - $(Fa & Ga & Fb & ~Gb) \lor (Fa & ~Ga & Fb & Gb)$
  - $(Fa & Ga & ~Fb & Gb) \lor (~Fa & Ga & Fb & Gb)$
  - $(Fa & Ga & ~Fb & ~Gb) \lor (~Fa & ~Ga & Fb & Gb)$
  - $(Fa & ~Ga & ~Fb & ~Gb) \lor (~Fa & ~Ga & Fb & ~Gb)$
  - $(~Fa & Ga & ~Fb & ~Gb) \lor (~Fa & ~Ga & Fb & Gb)$

- The measure $m^* [\therefore Pr^*]$ assigns equal probability to structure descriptions. $m^*$ takes individuals to be indistinguishable (analogy: Bose-Einstein statistics in QM), and $m^\dagger$ does not (analogy: Fermi-Dirac statistics in QM). Otherwise, they operate in the same way.
We can then define $\Pr^*(\bullet)$ and $\Pr^*(\bullet | \bullet)$ in terms of $m^*$, by assuming equiprobability of states within structure descriptions. Let’s compare the $m^\dagger$ and $m^*$ distributions over the 16 state descriptions of $L^{2,2}_Q$:

<table>
<thead>
<tr>
<th>$Fa$</th>
<th>$Ga$</th>
<th>$Fb$</th>
<th>$Gb$</th>
<th>State Descriptions ($s_i$)</th>
<th>$m^\dagger(s_i)$</th>
<th>$m^*(s_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>$Fa &amp; Ga &amp; Fb &amp; Gb$</td>
<td>1/16</td>
<td>1/10</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>$Fa &amp; Ga &amp; Fb &amp; \sim Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>$Fa &amp; Ga &amp; \sim Fb &amp; Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>$Fa &amp; Ga &amp; \sim Fb &amp; \sim Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>$Fa &amp; \sim Ga &amp; Fb &amp; Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>$Fa &amp; \sim Ga &amp; Fb &amp; \sim Gb$</td>
<td>1/16</td>
<td>1/10</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>$Fa &amp; \sim Ga &amp; \sim Fb &amp; Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>$Fa &amp; \sim Ga &amp; \sim Fb &amp; \sim Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>$\sim Fa &amp; Ga &amp; Fb &amp; Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>$\sim Fa &amp; Ga &amp; Fb &amp; \sim Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>$\sim Fa &amp; Ga &amp; \sim Fb &amp; Gb$</td>
<td>1/16</td>
<td>1/10</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>$\sim Fa &amp; Ga &amp; \sim Fb &amp; \sim Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>$\sim Fa &amp; \sim Ga &amp; Fb &amp; Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>$\sim Fa &amp; \sim Ga &amp; Fb &amp; \sim Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>$\sim Fa &amp; \sim Ga &amp; \sim Fb &amp; Gb$</td>
<td>1/16</td>
<td>1/20</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>$\sim Fa &amp; \sim Ga &amp; \sim Fb &amp; \sim Gb$</td>
<td>1/16</td>
<td>1/10</td>
</tr>
</tbody>
</table>

Now, $m^*$ does not have the “no learning from experience” property (*):
\[ \Pr^*(Fb \mid Fa) = \frac{\Pr^*(Fb \& Fa)}{\Pr^*(Fa)} = \frac{2 \cdot \frac{1}{10} + 2 \cdot \frac{1}{20}}{2 \cdot \frac{1}{10} + 6 \cdot \frac{1}{20}} = \frac{3}{5} > 2 \cdot \frac{1}{10} + 6 \cdot \frac{1}{20} = \frac{1}{2} = \Pr^*(Fb) \]

- Generally, \( \Pr^*(\bullet \mid \bullet) \) says that the more objects that are assumed to have \( F \), the more probable it is that other objects will also have \( F \). This is called *instantial relevance*. Carnap prefers \( \Pr^* \) over \( \Pr^\dagger \) for this reason.

- Is Carnap committed to the view that \( K_T \) favors certain state descriptions over others? He prefers \( m^* (Fa \& Ga \& Fb \& Gb) > m^* (Fa \& Ga \& Fb \& \sim Gb) \) to \( m^\dagger (Fa \& Ga \& Fb \& Gb) = m^\dagger (Fa \& Ga \& Fb \& \sim Gb) \). But, why?

- Carnap realizes that if a theory of "logical probability" is to provide a foundation for claims about *evidence*, then it *must* be able to furnish \( \Pr \)-models with *correlations* between *logically independent* claims.

- This is the tip of the iceberg for linguistic theories of "logical probability" like Wittgenstein’s and Carnap’s. Such theories are unable to emulate a broad enough range of probability functions, so as \( (D_3) \) to be generally *applicable* for modeling evidential relations in epistemic contexts.

- We will return to this *applicability* issue later in the course...
Carnapian Logical Probability: Analogy and Similarity 1

- Moreover, Pr* has other shortcomings. Carnap and others have endorsed the following “analogical” principle: a and b differing regarding F should weaken the inference from Ga to Gb, without completely undermining it:

\[
\text{Pr}(Gb \mid Ga) > \text{Pr}(Gb \mid Fa \& Ga \& \sim Fb) > \text{Pr}(Gb)
\]

Neither Pr* nor Pr† satisfies this condition. This will be a HW #3 problem.

- Generally: if a has properties \(P_1 \ldots P_n\), and b has \(P_1 \ldots P_{n-2}\), but lacks \(P_{n-1}\), that should be of some relevance to b’s having \(P_n\). In the \(n = 3\) case:

\[
\text{Pr}(Hb \mid Ha) > \text{Pr}(Hb \mid Fa \& Ga \& Ha \& Fb \& \sim Gb) > \text{Pr}(Hb \mid Fa \& Ga \& Ha \& \sim Fb \& \sim Gb) > \text{Pr}(Hb)
\]

- I.e., Differing on 2 properties should be worse than 1, but neither should completely undermine instantial relevance. Pr† and Pr* violate this.

- Note: this measures “similarity” by “shared predicate counting”. This is unfortunate, as it leads to Pr-functions that are language-relative…
Carnapian Logical Probability: Analogy and Similarity 2

- Let \(x_1, \ldots, x_n\) be the objects that fall under the predicate \(X\) \([i.e., \text{those in } \text{Ext}(X)]\). And, let \(s(x_1, \ldots, x_n)\) be some measure of “the degree to which the objects falling under \(X\) are similar to each other”.

- Carnap doesn’t offer much in the way of a theory of \(s(x_1, \ldots, x_n)\). But, his discussion suggests the following account of \(s(x_1, \ldots, x_n)\):

- Let \(P(x)\) be the set of predicates that \(x\) falls under. Then, define:

\[
s(x_1, \ldots, x_n) = \left| \bigcap_i P(x_i) \right|
\]

- That is, \(s(x_1, \ldots, x_n)\) is the size (cardinality) of the intersection of all the \(P(x_i)\). This is “the size of the set of shared predicates of the \(x_i\)”.

- There is a problem with this idea. Next, I will present an argument which shows that this measure of similarity is language variant.
Carnapian Logical Probability: Analogy and Similarity 3

• That is, “the degree of similarity of \( a \) and \( b \)” depends sensitively on the syntax of the language one uses to describe \( a \) and \( b \). Here’s why.

• The \( ABCD \) language consists of four predicates \( A, B, C, \) and \( D \). And, the \( XYZU \) language also has four predicates \( X, Y, Z, \) and \( U \) such that

\[
Xx \equiv A x \equiv B x, \quad Yx \equiv B x \equiv C x, \quad Zx \equiv A x, \quad \text{and} \quad Ux \equiv D x.
\]

\( ABCD \) and \( XYZU \) are (extra-systematically) expressively equivalent. Anything that can be said in \( ABCD \) can be said in \( XYZU \), and conversely — intuitively, there is no semantic difference between the two languages.

• Now, consider two objects \( a \) and \( b \) such that:

\[
Aa \& Ba \& Ca \& Da
\]

\[
A b \& \sim B b \& C b \& D b
\]

• Question: How similar are \( a \) and \( b \) in our “predicate-sharing” sense?
Carnapian Logical Probability: Analogy and Similarity 4

- Answer: That depends on which of our expressively equivalent languages we use to describe \( a \) and \( b \)! To see this, note that in \( XYZU \) we have:

\[
Xa \& Ya \& Za \& Ua \\
\sim Xb \& \sim Yb \& Zb \& Ub
\]

- Therefore, in \( ABCD \), \( a \) and \( b \) share three predicates. But, in \( XYZU \), \( a \) and \( b \) share only two predicates. Or, to use a modified notation, we have:

\[
s_{ABCD}(a, b) = 3 \neq 2 = s_{XYZU}(a, b)
\]

- Not good. Probabilities should not be language-variant. It shouldn’t matter which language you use to describe the world — semantically equivalent statements should be probabilistically indistinguishable.

- One consequence is that if \( p \models q \), \( x \models y \) and \( z \models u \), then we shouldn’t have both \( \Pr(p \mid x) > \Pr(p \mid z) \) and \( \Pr(q \mid u) > \Pr(q \mid y) \). Carnapian principles of analogy and similarity contradict this requirement.
Carnapian Logical Probability: Analogy and Similarity 5

- Here’s a concise way of stating the general Carnapian analogical principle:

\[(A) \text{ If } n > m, \text{ then } \Pr(Xa \mid Xb & s(a, b) = n) > \Pr(Xa \mid Xb & s(a, b) = m)\]

- Applying this principle to our example yields both of the following:

(1) \( \Pr(Da \mid Aa & Ba & Ca & Ab & \sim Bb & Cb) > \Pr(Da \mid Aa & Ba & Ca & Ab & \sim Bb & \sim Cb) \)

(2) \( \Pr(Ua \mid Xa & Ya & Za & \sim Xb & Yb & Zb) > \Pr(Ua \mid Xa & Ya & Za & \sim Xb & \sim Yb & Zb) \)

- Now, let \( p \equiv Da, q \equiv Ua, \) and

\( x \equiv Aa & Ba & Ca & Ab & \sim Bb & Cb, y \equiv Xa & Ya & Za & \sim Xb & \sim Yb & Zb, \)
\( z \equiv Xa & Ya & Za & \sim Xb & Yb & Zb, u \equiv Aa & Ba & Ca & Ab & \sim Bb & \sim Cb. \)

- Then, \( p \models q, x \models y \) and \( z \models u, \) but the Carnapian principle \((A)\) implies both (1) \( \Pr(p \mid x) > \Pr(p \mid u), \) and (2) \( \Pr(q \mid z) > \Pr(q \mid y). \) Bad.

- It seems that principle \((A)\) must go. Otherwise, some restriction on the choice of language is required to block inferring both (1) and (2) from it.