Philosophy 148 — Announcements & Such

- The mid-term is Thursday. I discussed it last night. These are fair game:
  - True/False/short calculation (quiz-like) questions on formal stuff.
  - Short answer questions about truth, and/or interpretations of probability, and/or the lottery paradox, and/or the preface paradox.
  - A question about Dutch Books. [Joyce's argument is *not* fair game.]
- I've posted my LaTeX source for the HW #2 hints handout (see website).
- Today's Agenda:
  - A brief review of the preface paradox and the consistency norm.
  - Joyce's "Gradational Accuracy" argument for Epistemic Probabilism
    * The set-up of Joyce's framework
    * The premises of his argument
    * Questioning the premises (different measures of "accuracy")
    * Joyce's response to these worries
- Next: "Logical Probability" and then Confirmation Theory.

The Preface Paradox and Epistemic Rationality – Review

- Last time, I gave an argument to the effect that (almost) everyone — on reflection — has inconsistent beliefs, and that this is epistemically reasonable. The argument runs as follows. If it goes wrong, where?
  - *A fortiori*, you believe each element of your belief set $\mathcal{B}$.
  - You know your belief set $\mathcal{B}$ is very large and complex.
  - You know that all human beings are prone to epistemic error, and $\therefore$ (reasonably) believe that all large belief sets contain some falsehoods.
  - On this basis, you (reasonably) believe $\mathcal{B}$ contains some falsehoods.
  - But, now, $\mathcal{B}$ is *inconsistent*, and you $\therefore$ violate the consistency norm!
- This argument calls into question the consistency norm for belief. Maybe there is *some* sense in which inconsistency is "epistemically bad", but then what's wrong with someone whose $\mathcal{B}$ satisfies the above premises?
- In any case, this argument should at least have you worried about the consistency norm. Next, we move on to "epistemic probabilism".

Epistemic Probability: Joyce's Argument for Epistemic Probabilism I

- Joyce's argument is based on the following assumptions and set-up:
  - Each agent S has a finite (or countable) set of propositions $\mathcal{B}$ (it need not be a Boolean algebra) that they are capable of entertaining and assigning some degree of credence (*i.e.*, epistemically rational d.o.b) to.
  - We are concerned with the *accuracy* of an agent's credence function q. We use the notation $I(q, w)$ to denote the *in*accuracy of q in world w.
  - $I(q, w) = \sum_{X \in \mathcal{B}} \rho(q(X), w(X))$, where $q(X)$ is S's credence in X, $w(X)$ is the truth-value of X at w (0 if X is false in w, 1 if X is true in w), and $\rho(q(X), w(X))$ is a measure of "distance" between $q(X)$ and $w(X)$.
  - Note two things about I. First, I is defined *piecewise*. Second, I is an *additive* function of the "distances from the truth" of each $p \in \mathcal{B}$.
  - The **Norm of Gradational Accuracy** (NGA) says that an epistemically rational agent strives to have credences that are *as close to the truth as possible*, *i.e.*, to *minimize* $I(q, w)$ in some suitable sense of "minimize".

Epistemic Probability: Joyce's Argument for Epistemic Probabilism II

- **Theorem.** Given four assumptions $\mathcal{A}$ (to be discussed below) about I, if q is not a probability function on $\mathcal{B}$, then there exists a probabilistic credence function q', which is strictly more accurate than q, in every possible world w (*i.e.*, *no matter what the truth-values of the p's in* $\mathcal{B}$ *turn out to be*). *I.e.*, $\exists$ a Pr-function q' such that $I(q', w) < I(q, w)$, for all w.
- This is a nice *epistemic* result. It purports to show that non-probabilistic credence functions are *epistemically* inferior to probabilistic ones — in terms of their *accuracy*. Of course, the devil is in the details: $\mathcal{A}$.
- Next, I'll discuss each of the assumptions in $\mathcal{A}$ (*i.e.*, the pre-conditions of the above Theorem). As we will see, the first two of these seem very plausble. The second pair of assumptions is more controversial.
- We will discuss at some length some objections to the second pair advanced by Patrick Maher. One assumption in the first pair also been called into question by Gibbard (see his paper). We won't discuss that.

1. **Dominance.** If $q_1(y) = q_2(y)$, for all $y \neq x$, then $I(q_1, w) > I(q_2, w)$ iff $\rho(q_1(x), w(x)) > \rho(q_2(x), w(x))$. [Nobody objects to this one!]
   – This ensures the appropriateness of the linear, piecewise way of thinking about $I(q, w) = \sum_{X \in \mathcal{B}} \rho(q(X), w(X))$. *No Holism*!
2. **Normality**. If, for all x, $|w(x) - q_1(x)| = |w'(x) - q_2(x)|$, then $I(q_1, w) = I(q_2, w')$. [Allan Gibbard objects to this. We won't discuss.]
   – *Over*estimating a proposition's p's truth-value by ϵ is counted as exactly as inaccurate as *under*estimating its truth-value by ϵ.

---

3. **Weak Convexity** (WC). If $I(q_1, w) = I(q_2, w)$, then $I(q_1, w) \geq I(\frac{1}{2}q_1 + \frac{1}{2}q_2, w)$ with identity *only if* $q_1 = q_2$.
   – *Averaging* two equally inaccurate q's [q_1, q_2] can't yield a *strictly more inaccurate* $\bar{q}$. And, $\{\bar{q}, q_1, q_2\}$ are *equally inaccurate* ***only if*** $q_1 = q_2$.
4. **Symmetry** (S). If $I(q_1, w) = I(q_2, w)$ then, for any $\lambda \in [0, 1]$, $I(\lambda \cdot q_1 + (1 - \lambda) \cdot q_2, w) = I((1 - \lambda) \cdot q_1 + \lambda \cdot q_2, w)$.
   – Symmetric weighted averages of equally inaccurate q's remain so.

- (WC) and (S) are the really controversial (and non-obvious) ones. To wit ...

Epistemic Probability: Joyce's Argument for Epistemic Probabilism III

- (WC) and (S) place constraints on which underlying distance measures ρ are admissible. There are *many* measures of "distance" ρ which lead to inaccuracies I satisfying (WC) & (S). *E.g.*, the *square* difference measure $\rho^{\dagger}$:

$$\rho^{\dagger}(q(X), w(X)) \stackrel{\text{def}}{=} (q(X) - w(X))^2$$

  leads to an I satisfying (WC) and (S). Distance measures like $\rho^{\dagger}$ are called *proper scoring rules* (PSRs) [see below]. Are all measures of distance PSRs?

- Patrick Maher points out that the *absolute value* difference measure ρ^*:

$$\rho^*(q(X), w(X)) \stackrel{\text{def}}{=} |q(X) - w(X)|$$

  leads to an I which *violates* (WC) & (S). And, Maher shows that ρ^* can*not* undergird Joyce's Theorem, since ρ^* allows non-probabilistic credence functions to be more accurate than probabilistic ones (at some w's)!

- Maher argues that ρ^* is an adequate measure of distance. If Maher is right, then this is a serious problem for any approach like Joyce's.

Epistemic Probability: Joyce's Argument for Epistemic Probabilism IV

- ρ^* leads to an I that violates both (WC) and (S). To find a counterexample to (WC) for ρ^*, one needs at least 2 propositions in $\mathcal{B}$ (HW!). Here's one:
  - Assume A is true, and B is false, in w. So, $w(A) = 1$ and $w(B) = 0$.
  - Let $q_1(A) = q_1(B) = 1$, and $q_2(A) = q_2(B) = 0$.
  - Therefore, $q_3(A) = \dfrac{q_1(A) + q_2(A)}{2} = \dfrac{1+0}{2} = \dfrac{1}{2}$.
  - And, $q_3(B) = \dfrac{q_1(B) + q_2(B)}{2} = \dfrac{1+0}{2} = \dfrac{1}{2}$.
  - Thus, $I(q_1, w) = \rho^*(q_1(A), w) + \rho^*(q_1(B), w) = |1-1| + |1-0| = 1$.
  - And, $I(q_2, w) = \rho^*(q_2(A), w) + \rho^*(q_2(B), w) = |1-0| + |0-0| = 1$.
  - And, $I(q_3, w) = \rho^*(q_3(A), w) + \rho^*(q_3(B), w) = |\frac{1}{2} - 1| + |\frac{1}{2} - 0| = 1$.
  - So, we have $I(q_1, w) = I(q_2, w) = I(q_3, w)$, but $q_1 \neq q_2$, violating (WC).
  - Note: $\rho^\dagger(q_3(A), w) = \rho^\dagger(q_3(B), w) = (\frac{1}{2} - 1)^2 = \frac{1}{4}$, which blocks the counterexample. [HW: there are *no* 2-proposition counterexamples for $\rho^\dagger$.]

Epistemic Probability: Joyce's Argument for Epistemic Probabilism V

- To find a counterexample to (S) for ρ^*, one also needs ≥ 2 propositions:
  - Assume A is true, and B is true, in w. So, $w(A) = 1$ and $w(B) = 1$.
  - Let $q_1(A) = 1$ and $q_1(B) = 3$. And, let $q_2(A) = q_2(B) = 0$.
  - Thus, $I(q_1, w) = \rho^*(q_1(A), w) + \rho^*(q_1(B), w) = |1 - 1| + |1 - 3| = 2$.
  - And, $I(q_2, w) = \rho^*(q_2(A), w) + \rho^*(q_2(B), w) = |1 - 0| + |1 - 0| = 2$.
  - $I(q_1, w) = I(q_2, w)$. So, (S) $\Rightarrow I(q_3, w) = I(q_4, w)$, where $\lambda = \frac{1}{4}$; and:

$$q_3 = \lambda \cdot q_1 + (1 - \lambda) \cdot q_2 = \tfrac{1}{4} \cdot q_1 + \tfrac{3}{4} \cdot q_2$$
$$q_4 = (1 - \lambda) \cdot q_1 + \lambda \cdot q_2 = \tfrac{3}{4} \cdot q_1 + \tfrac{1}{4} \cdot q_2$$

But,

$$\begin{aligned} I(q_3, w) &= \rho^*(q_3(A), w) + \rho^*(q_3(B), w) \\ &= \left|1 - \left(\frac{1}{4} \cdot 1 + \frac{3}{4} \cdot 0\right)\right| + \left|1 - \left(\frac{1}{4} \cdot 3 + \frac{3}{4} \cdot 0\right)\right| = \frac{3}{4} + \frac{1}{4} = 1 \\ I(q_4, w) &= \rho^*(q_4(A), w) + \rho^*(q_4(B), w) \\ &= \left|1 - \left(\frac{3}{4} \cdot 1 + \frac{1}{4} \cdot 0\right)\right| + \left|1 - \left(\frac{3}{4} \cdot 3 + \frac{1}{4} \cdot 0\right)\right| = \frac{1}{4} + \frac{5}{4} = \frac{3}{2} \end{aligned}$$

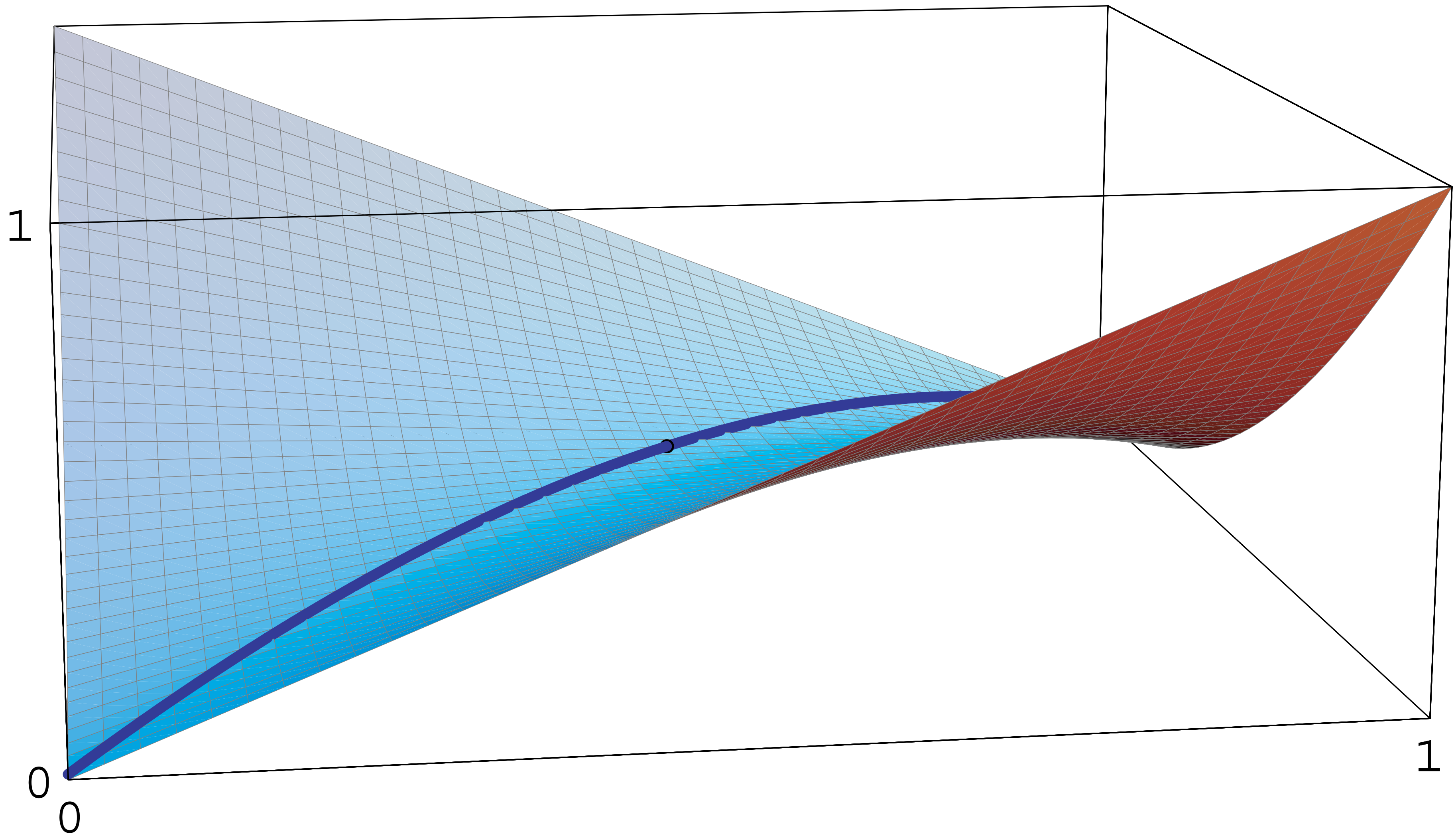
Epistemic Probability: Joyce's Argument for Epistemic Probabilism VI

- **Joyce's Reply**: define *expected distance of* q' *(according to* q*) from the truth concerning* X*, relative to* ρ*,* $EDT(q', q, X, \rho)$, in the following way:

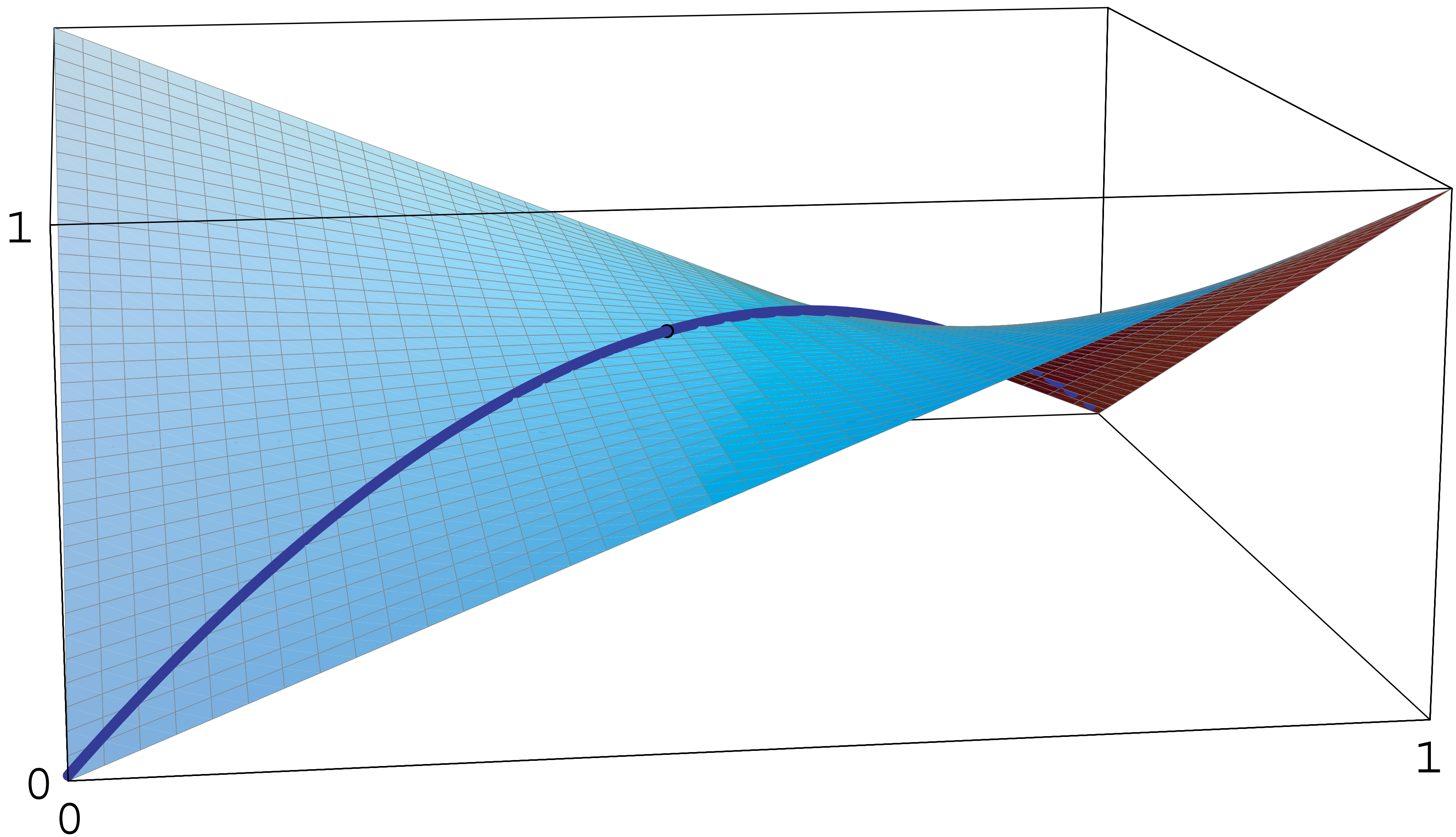
$$\begin{aligned} EDT(q', q, X, \rho) &= q(X) \cdot \rho(q'(X), 1) + (1 - q(X)) \cdot \rho(q'(X), 0) \\ &= q \cdot \rho(q', 1) + (1 - q) \cdot \rho(q', 0) \end{aligned}$$

- **Theorem.** If $\rho = \rho^*$, then $EDT(q', q, X, \rho)$ can be minimized *only* by assigning $q'(X) = 0$, $q'(X) = \frac{1}{2}$, or $q'(X) = 1$. If $\rho = \rho^\dagger$, $EDT(q', q, X, \rho)$ will be minimized when $q'(X) = q(X)$, which can lie *anywhere* on $[0, 1]$.
- If you aim to minimize your *expected inaccuracy* of your own q, then you shouldn't use ρ^*, since this will lead you to view your current degrees of belief as inferior to one that assigns 0, 1 or $\frac{1}{2}$ to *every* X in $\mathcal{B}$.
- But, it's OK to use $\rho^\dagger$, since this leads to a "*stable*" q. See plots, below.
- Questions: *why* minimize $\boldsymbol{EDT}$? Wasn't the goal to minimize DT? Also, why do we get to *assume* here that $q(\sim X) = 1 - q(X)$? Question begging?

Plot of $EDT(q', q, X, \rho^{\dagger})$ — the curve $q' = q$ is a *stable* "saddle".



Plot of $EDT(q', q, X, \rho^*) - q' = q$ is *un*stable (except at $\{0, \frac{1}{2}, 1\}$).



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 ρ 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 $\mathcal{B}$.
- 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 ρ^{*} violates two of Joyce's axioms: (WC) and (S), and he argues that ρ^{*} is nonetheless a reasonable measure of "distance from truth". Maher also shows that Joyce's Theorem *fails* if one uses ρ^{*}.
- Joyce's response appeals to "Expected Distance from Truth". But, this undermines one of the nice things about Joyce's original argument — that it made no appeal to Expected Utility Theory (which seems *pragmatic*).

Whither Logical Probability? 1

- 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 ($\vDash$).
- Inductive Logic aims to explicate a quantitative measure $\mathfrak{c}(C, P)$ 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 $\mathfrak{c}$ should quantitatively generalize deductive entailment:

($\mathcal{D}_1$) The relations of deductive entailment and deductive refutation should be captured as limiting (extreme) values of $\mathfrak{c}$ with cases of 'partial entailment' and 'partial refutation' lying somewhere on a $\mathfrak{c}$-continuum (or range) between these extreme values of the confirmation function.

- Of course, *conditional probability* $\mathfrak{c}(C, P) = \Pr(C \mid P)$ satisfies $\mathcal{D}_1$.

Whither Logical Probability? 2

- Another key historical desideratum for inductive logic is:

($\mathcal{D}_2$) 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 [$\mathfrak{c}$] which is likewise objective and logical, *viz.*, . . . degree of confirmation."

- Carnap on the terms "logical" and "objective" as they apply to $\mathfrak{c}$.

  "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 $\mathfrak{c}$ is objective means this: if a certain $\mathfrak{c}$ value holds for a hypothesis (H) with respect to evidence (E), then the value $\mathfrak{c}(H, E)$ is independent of what any person may happen to think about these sentences."

Whither Logical Probability? 3

- While $\Pr(C \mid P)$ clearly satisfies $\mathcal{D}_1$, it is unclear whether $\Pr(C \mid P)$ satisfies $\mathcal{D}_2$. If $\mathfrak{c}(C, P) \overset{\text{def}}{=} \Pr(C \mid P)$, then, we seem to need $\Pr(C \mid P)$ *itself* to be logical if $\mathfrak{c}(C, P)$ 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* ($\mathfrak{m}$). 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 $\mathfrak{m}$ 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 $\mathfrak{m}$ would assign $\mathfrak{m}(s_i) = \frac{1}{2^n}$, for all state descriptions s_i of $\mathcal{L}_P^n$. Here's what $\mathfrak{m}$ looks like over an $\mathcal{L}_P^3$ language (this should look familiar!):

A	B	C	State Descriptions	$\mathfrak{m}(s_i)$
T	T	T	$s_1 = A \,\&\, B \,\&\, C$	$\mathfrak{m}(s_1) = 1/8$
T	T	F	$s_2 = A \,\&\, B \,\&\, {\sim}C$	$\mathfrak{m}(s_2) = 1/8$
T	F	T	$s_3 = A \,\&\, {\sim}B \,\&\, C$	$\mathfrak{m}(s_3) = 1/8$
T	F	F	$s_4 = A \,\&\, {\sim}B \,\&\, {\sim}C$	$\mathfrak{m}(s_4) = 1/8$
F	T	T	$s_5 = {\sim}A \,\&\, B \,\&\, C$	$\mathfrak{m}(s_5) = 1/8$
F	T	F	$s_6 = {\sim}A \,\&\, B \,\&\, {\sim}C$	$\mathfrak{m}(s_6) = 1/8$
F	F	T	$s_7 = {\sim}A \,\&\, {\sim}B \,\&\, C$	$\mathfrak{m}(s_7) = 1/8$
F	F	F	$s_8 = {\sim}A \,\&\, {\sim}B \,\&\, {\sim}C$	$\mathfrak{m}(s_8) = 1/8$

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* $p \vDash q$ or $\sim p \vDash q$ or $q \vDash p$ or $\sim q \vDash p$). [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 $p \vDash q$. [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? Two approaches: (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 background knowledge 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:

$$\Pr(s_i \mid K) = \Pr(s_j \mid K), \text{ for all } i, j.$$

- 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 *de*ductive logic, *epistemic* principles seem to make for odd *logical* constraints. Why should *in*ductive logic be any different?

- Presumably, the (PI) approach goes as follows. Assume that "all we know is logic". Call this background knowledge $K_\top$. Assume that $K_\top$ does not favor any state description over any other. Then, $\mathfrak{m}$ can be thought of as $\Pr(\bullet \mid K_\top)$, and we have $\Pr(s_i \mid K_\top) = \Pr(s_j \mid K_\top)$, 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.*, $\mathfrak{c}(s_i, K) = \mathfrak{c}(s_j, K)$.
  - If we *assume* that $\mathfrak{c}(x, y) = \Pr(x \mid y)$, then *logic alone* implies that:
  $$\mathfrak{c}(s_i, K) = \mathfrak{c}(s_j, K) \Rightarrow \Pr(s_i \mid K) = \Pr(s_j \mid K), \text{ for } \textit{all } i \text{ and } j.$$
  - $\therefore$ *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. $\square$

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.*,

    $$\mathfrak{c}(x, y) = \Pr(x \mid y)$$

  - If it were obvious that $\Pr(x \mid y)$ is the best (probabilistic) explication of $\mathfrak{c}(x, y)$, 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 $\mathfrak{c}(x, y)$.
- 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, ...) and m constants (a, b, ...).
- 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 $\mathcal{L}_Q^{2,2}$:

$Fa \,\&\, Ga \,\&\, Fb \,\&\, Gb$	$Fa \,\&\, Ga \,\&\, Fb \,\&\, {\sim}Gb$	$Fa \,\&\, Ga \,\&\, {\sim}Fb \,\&\, Gb$
$Fa \,\&\, Ga \,\&\, {\sim}Fb \,\&\, {\sim}Gb$	$Fa \,\&\, {\sim}Ga \,\&\, Fb \,\&\, Gb$	$Fa \,\&\, {\sim}Ga \,\&\, Fb \,\&\, {\sim}Gb$
$Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, Gb$	$Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, {\sim}Gb$	${\sim}Fa \,\&\, Ga \,\&\, Fb \,\&\, Gb$
${\sim}Fa \,\&\, Ga \,\&\, Fb \,\&\, {\sim}Gb$	${\sim}Fa \,\&\, Ga \,\&\, {\sim}Fb \,\&\, Gb$	${\sim}Fa \,\&\, Ga \,\&\, {\sim}Fb \,\&\, {\sim}Gb$
${\sim}Fa \,\&\, {\sim}Ga \,\&\, Fb \,\&\, Gb$	${\sim}Fa \,\&\, {\sim}Ga \,\&\, Fb \,\&\, {\sim}Gb$	${\sim}Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, Gb$
	${\sim}Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, {\sim}Gb$	

Carnapian Monadic Predicate Logical Probability 2

- Following (PI), Carnap's first measure function $\mathfrak{m}^\dagger$ assigns *equal probability* to each state description s_i of $\mathcal{L}_Q^{m,n}$. In our $\mathcal{L}_Q^{2,2}$, $\mathfrak{m}^\dagger(s_i) = \frac{1}{16}$.
- We extend $\mathfrak{m}^\dagger$ 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 $\ulcorner \Pr^\dagger(q \mid p) = x \urcorner$ 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 $\mathfrak{m}^\dagger$ *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 *rejects* the measure function $\mathfrak{m}^\dagger$ in favor of an alternative measure function $\mathfrak{m}^*$, for epistemic-sounding reasons.
- Carnap notes that $\mathfrak{m}^\dagger$ causes $\Pr^\dagger$ to have the following property ($b \neq a$):

$$(*) \qquad \Pr^\dagger(Fb \mid Fa) = \frac{\Pr^\dagger(Fb \,\&\, Fa)}{\Pr^\dagger(Fa)} = \frac{4 \cdot \frac{1}{16}}{8 \cdot \frac{1}{16}} = \frac{1}{2} = 8 \cdot \frac{1}{16} = \Pr^\dagger(Fb)$$

- 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 *generalizes* to *any* number of Fs: $\Pr^\dagger(Fa_1 \mid Fa_2 \,\&\, \cdots \,\&\, Fa_m) = \Pr^\dagger(Fa_1)$.
- Carnap (*et al*) characterize this (in *epistemic* terms) as $\mathfrak{m}^\dagger$ leading to a function $\Pr^\dagger$ that fails to allow for "learning from experience". This is *epistemic and diachronic*, since it assumes *learning-by-conditionalization*.
- Carnap views this consequence of applying (PI) to the state descriptions of $\mathcal{L}_Q^{m,n}$ as unacceptable. Does he then think that $K_\top$ "favors" some state descriptions over others. But, which ones? Enter the $\mathfrak{m}^*$ measure ...

Carnapian Monadic Predicate Logical Probability 4

- Two state descriptions s_i and s_j in $\mathcal{L}_Q^{m,n}$ are *permutations* of each other if one can be obtained from the other by a permutation of constants.
- "$Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, Gb$" can be obtained from "${\sim}Fa \,\&\, Ga \,\&\, Fb \,\&\, {\sim}Gb$" by permuting "$a$" and "$b$". Thus, "$Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, Gb$" and "${\sim}Fa \,\&\, Ga \,\&\, Fb \,\&\, {\sim}Gb$" are permutations of each other (in $\mathcal{L}_Q^{2,2}$).
- A *structure description* is a disjunction of state descriptions, each of which is a permutation of the others. $\mathcal{L}_Q^{2,2}$ has 10 structure descriptions:

$(Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, Gb) \vee ({\sim}Fa \,\&\, Ga \,\&\, Fb \,\&\, {\sim}Gb)$
$(Fa \,\&\, Ga \,\&\, Fb \,\&\, {\sim}Gb) \vee (Fa \,\&\, {\sim}Ga \,\&\, Fb \,\&\, Gb)$
$(Fa \,\&\, Ga \,\&\, {\sim}Fb \,\&\, Gb) \vee ({\sim}Fa \,\&\, Ga \,\&\, Fb \,\&\, Gb)$
$(Fa \,\&\, Ga \,\&\, {\sim}Fb \,\&\, {\sim}Gb) \vee ({\sim}Fa \,\&\, {\sim}Ga \,\&\, Fb \,\&\, Gb)$
$(Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, {\sim}Gb) \vee ({\sim}Fa \,\&\, {\sim}Ga \,\&\, Fb \,\&\, {\sim}Gb)$
$({\sim}Fa \,\&\, Ga \,\&\, {\sim}Fb \,\&\, {\sim}Gb) \vee ({\sim}Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, Gb)$

$Fa \,\&\, {\sim}Ga \,\&\, Fb \,\&\, {\sim}Gb$
$Fa \,\&\, Ga \,\&\, Fb \,\&\, Gb$
${\sim}Fa \,\&\, Ga \,\&\, {\sim}Fb \,\&\, Gb$
${\sim}Fa \,\&\, {\sim}Ga \,\&\, {\sim}Fb \,\&\, {\sim}Gb$

- The measure $\mathfrak{m}^*$ [$\therefore$ $\Pr^*$] assigns equal probability to *structure descriptions*. $\mathfrak{m}^*$ takes individuals to be *indistinguishable* (analogy: Bose-Einsten statistics in QM), and $\mathfrak{m}^\dagger$ does not (analogy: Fermi-Dirac statistics in QM). Otherwise, they operate in the same way.

Carnapian Monadic Predicate Logical Probability 5

- We can then define $\Pr^*(\bullet)$ and $\Pr^*(\bullet \mid \bullet)$ in terms of $\mathfrak{m}^*$, by assuming equiprobability of *states within* structure descriptions. Let's compare the $\mathfrak{m}^\dagger$ and $\mathfrak{m}^*$ distributions over the 16 state descriptions of $\mathcal{L}_Q^{2,2}$:

Fa	Ga	Fb	Gb	State Descriptions (s_i)	$\mathfrak{m}^\dagger(s_i)$	$\mathfrak{m}^*(s_i)$
T	T	T	T	$Fa \& Ga \& Fb \& Gb$	1/16	1/10
T	T	T	F	$Fa \& Ga \& Fb \& {\sim}Gb$	1/16	1/20
T	T	F	T	$Fa \& Ga \& {\sim}Fb \& Gb$	1/16	1/20
T	T	F	F	$Fa \& Ga \& {\sim}Fb \& {\sim}Gb$	1/16	1/20
T	F	T	T	$Fa \& {\sim}Ga \& Fb \& Gb$	1/16	1/20
T	F	T	F	$Fa \& {\sim}Ga \& Fb \& {\sim}Gb$	1/16	1/10
T	F	F	T	$Fa \& {\sim}Ga \& {\sim}Fb \& Gb$	1/16	1/20
T	F	F	F	$Fa \& {\sim}Ga \& {\sim}Fb \& {\sim}Gb$	1/16	1/20
F	T	T	T	${\sim}Fa \& Ga \& Fb \& Gb$	1/16	1/20
F	T	T	F	${\sim}Fa \& Ga \& Fb \& {\sim}Gb$	1/16	1/20
F	T	F	T	${\sim}Fa \& Ga \& {\sim}Fb \& Gb$	1/16	1/10
F	T	F	F	${\sim}Fa \& Ga \& {\sim}Fb \& {\sim}Gb$	1/16	1/20
F	F	T	T	${\sim}Fa \& {\sim}Ga \& Fb \& Gb$	1/16	1/20
F	F	T	F	${\sim}Fa \& {\sim}Ga \& Fb \& {\sim}Gb$	1/16	1/20
F	F	F	T	${\sim}Fa \& {\sim}Ga \& {\sim}Fb \& Gb$	1/16	1/20
F	F	F	F	${\sim}Fa \& {\sim}Ga \& {\sim}Fb \& {\sim}Gb$	1/16	1/10

- 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 $K_\top$ favors certain state descriptions over others? He prefers $\mathfrak{m}^*(Fa \,\&\, Ga \,\&\, Fb \,\&\, Gb) > \mathfrak{m}^*(Fa \,\&\, Ga \,\&\, Fb \,\&\, {\sim}Gb)$ to $\mathfrak{m}^\dagger(Fa \,\&\, Ga \,\&\, Fb \,\&\, Gb) = \mathfrak{m}^\dagger(Fa \,\&\, Ga \,\&\, Fb \,\&\, {\sim}Gb)$. 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 a much bigger 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 to be generally applicable to intuitive models of evidential relations.
- We will return to this issue later in the course …

Carnapian Monadic Predicate Logical Probability 6

- Moreover, $\Pr^*$ also has 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*:

$$\Pr(Gb \mid Ga) > \Pr(Gb \mid Fa \,\&\, Ga \,\&\, {\sim}Fb) > \Pr(Gb)$$

Neither $\Pr^*$ nor $\Pr^\dagger$ 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:

$$\Pr(Hb \mid Ha) > \Pr(Hb \mid Fa \,\&\, Ga \,\&\, Ha \,\&\, Fb \,\&\, {\sim}Gb)$$

$$> \Pr(Hb \mid Fa \,\&\, Ga \,\&\, Ha \,\&\, {\sim}Fb \,\&\, {\sim}Gb) > \Pr(Hb)$$

- *I.e.*, Differing on 2 properties should be worse than 1, but neither should completely undermine instantial relevance. $\Pr^\dagger$ 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*. . .