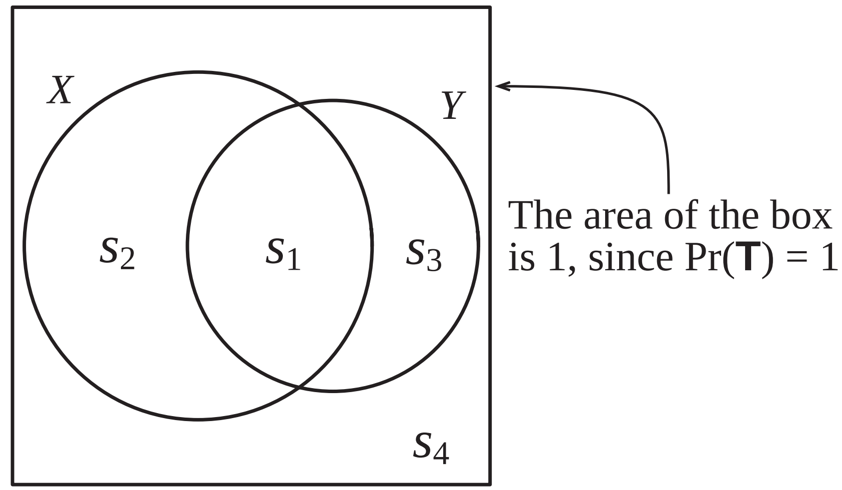
Announcements & Overview

- Administrative Stuff
  - **HW #3 grades & solutions have been posted**
  - **The mid-term grades have also been posted**
    - The class did very well, generally.
    - If you'd like to pick up your exam, please stop by my office hours sometime (they are in my office).
  - **HW #4 has been posted — due this Friday (March 25)**
    - This one consists of six (6) validity testing problems. The last 3 *require* the "short method" (either method is OK for the first 3).
  - **Consult my "Short Method" handout for detailed examples of the "short method" and its presentation (to be discussed today).**
  - **I will not be holding office hours today.**
- Unit #4 — *Probability & Inductive Logic*

Unit #4: Probability & Inductive Logic

- In this unit, we will build on the concepts of Unit #3, by introducting *probabilities* over LSL sentences. This will only require some simple high-school algebra, over-and-above LSL truth-table methods.
- Intuitively, one can think of "the probability that p is true" as "the *proportion* of possible worlds in which p is true." We will use the notation ⌜$\Pr(p)$⌝ to abbreviate ⌜the probability that p is true⌝.
- One informal way to picture Pr's is to use what I like to call *Stochstic Venn Diagrams* (SVDs), which use *areas* to represent *probabilities*. That is, the area of a region in an SVD is proportional to its probability.
- A more general way to visualize probabilities (*i.e.*, probability *distributions*) is to use what I call *Stochastic Truth Tables* (STTs), which have a column for the probabilities assigned to each state.
- Example: here is a (numerical) probability distribution over the states induced by a language with two atomic sentences $\{X, Y\}$.

X	Y	States	$\Pr(s_i)$
$\top$	$\top$	s_1	$\frac{4}{24} \approx 0.166$
$\top$	$\bot$	s_2	$\frac{6}{24} = 0.25$
$\bot$	$\top$	s_3	$\frac{3}{24} = 0.125$
$\bot$	$\bot$	s_4	$\frac{11}{24} \approx 0.458$



- We can make this example more concrete by imagining that we have an urn containing 24 objects, which are either black (X) or white ($\sim X$) and either metallic (Y) or plastic ($\sim Y$). The distribution of properties is then given by the proportions listed in the last column of the STT on the left.
- Suppose we are going to sample an object o (at random) from the urn. Then, *e.g.*, *the probability that o is black and metallic* is given by the proportion of objects in the urn which satisfy the state description $X \,\&\, Y$.
- The SVD on the right is drawn to scale, in the sense that the area of each region (corresponding to each state) is proportional to its probability.

- Once we've introduced the general theory of probability (over sentential/LSL languages), we'll use that theory to give an account of (inductive) *argument strength* that is more general than validity.

- Some invalid arguments seem (intuitively) logically *stronger than* others:

$$(1) \quad \begin{array}{c} P \vee Q \\ \therefore P \end{array} \qquad\qquad (2) \quad \begin{array}{c} P \vee \sim P \\ \therefore P \end{array}$$

- *Inductive* logic should *theoretically ground* our intuition that (1) is a *logically stronger* argument than (2) is. Note: *neither* argument is *valid*.

- More ambitiously, an inductive logician might aim for a theory of "the *degree* to which the premises of an argument *confirm* its conclusion".

- This ambitious project would aim to characterize a *function* $\mathfrak{c}(C, \mathcal{P})$. And, an intuitive requirement would be that this function be such that:

$$\mathfrak{c}(P, P \vee Q) > \mathfrak{c}(P, P \vee \sim P).$$

- In Unit #4, we will explain how *probability calculus* can be used to explicate these sorts of "confirmation/support functions" $\mathfrak{c}$.

Probability Calculus I

- The numerical probability distribution above (involving sampling from an urn) is just a special case. The *probability calculus* is a *general theory* of probability (over sentential/LSL languages).
- The two ingredients of probability calculus are as follows:
  - Truth-functional logic (*i.e.*, LSL and truth-table methods).
  - High School Algebra: basic algebraic operations over real numbers and variables ranging over real numbers. This includes equations and inequalities and some simple facts involving polynomials.
- We've already covered truth-functional logic, in Unit #3.
- The "high-school algebra" part of probability calculus is just a fragment of what I'm sure you all learned in high-school about (real) algebra.
- We'll begin with some (abstract) definitions and assumptions.

Probability Calculus II

X	Y	Z	States	$\Pr(s_i)$
$\top$	$\top$	$\top$	$s_1 = X \,\&\, Y \,\&\, Z$	a_1
$\top$	$\top$	$\bot$	$s_2 = X \,\&\, Y \,\&\, {\sim}Z$	a_2
$\top$	$\bot$	$\top$	$s_3 = X \,\&\, {\sim}Y \,\&\, Z$	a_3
$\top$	$\bot$	$\bot$	$s_4 = X \,\&\, {\sim}Y \,\&\, {\sim}Z$	a_4
$\bot$	$\top$	$\top$	$s_5 = {\sim}X \,\&\, Y \,\&\, Z$	a_5
$\bot$	$\top$	$\bot$	$s_6 = {\sim}X \,\&\, Y \,\&\, {\sim}Z$	a_6
$\bot$	$\bot$	$\top$	$s_7 = {\sim}X \,\&\, {\sim}Y \,\&\, Z$	a_7
$\bot$	$\bot$	$\bot$	$s_8 = {\sim}X \,\&\, {\sim}Y \,\&\, {\sim}Z$	a_8

- Each state s_i has an associated probability $\Pr(s_i) = a_i$.
- The probability of any (non-contradictory) LSL statement p can be calculated *via* its *disjunctive normal form*. That is, we have the following general definition (note: contradictions have probability *zero*).

$$\Pr(p) \overset{\text{def}}{=} \sum_{s_i \vDash p} \Pr(s_i)$$

- For instance, returning to our DNF example from above, we have:

$$\Pr(X \,\&\, (Y \leftrightarrow Z)) = \Pr(s_1 \vee s_4) = a_1 + a_4$$

- There are just *two* general constraints on the (basic) probabilities a_i.

Constraint #1. Each a_i is *on the unit interval* $[0,1]$.

∗ Intuitively, $\Pr(p) = 0$ means p has "0% chance" of occurring, and $\Pr(p) = 1$ means p has "100% chance" of occurring. And, *all* probabilities must fall somewhere between these two extremes.

Constraint #2. The a_i must *add up to 1* (*i.e.*, $\sum_{i=1}^{2^n} a_i = 1$).

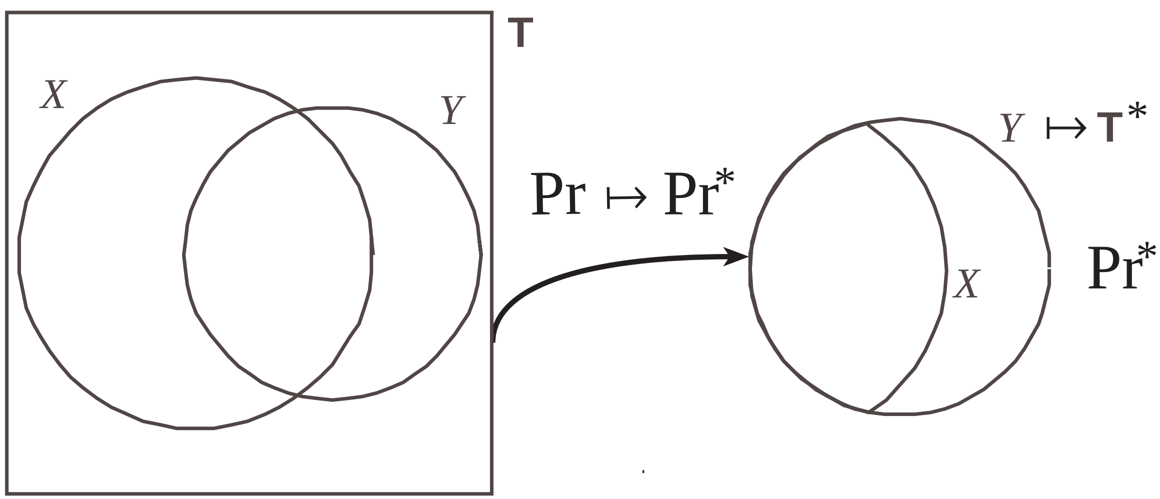
∗ Exactly one of the 2^n states must obtain, so when we add up all of the probabilities of all of the states, we get the maximal value of 1. Another way to see this is to note that the disjunction of all the states is a *tautology*, which must have *maximal* probability.

☞ With these three simple stipulations in hand, we can now derive *any* truth regarding probability calculus from simple high-school algebra.

Probability Calculus III

- Our SVD visualization of probability functions allows us to better understand the motivation for our definition of *conditional probability*.
- Intuitively, $\Pr(X \mid Y)$ is supposed to be the probability of X *given that* Y *is true*. So, when we conditionalize on Y, it's like *supposing* Y *to be true*.
- If we suppose Y to be true, then this is like *treating the* Y*-circle as if it is the entire bounding box of a (new, "conditionalized") Venn Diagram*.
- This is like *moving to a new* $\Pr^*$*-function, according to which* $\Pr^*(Y) = 1$.

$\Pr(X)$



$$\Pr^*(X) = \Pr(X \mid Y) = \frac{\Pr(X \,\&\, Y)}{\Pr(Y)}$$

- And, this is precisely how conditional probability is defined.

  Def. of Conditional Probability. $\Pr(p \mid q) \overset{\text{def}}{=} \dfrac{\Pr(p \,\&\, q)}{\Pr(q)}$, *if* $\Pr(q) > 0$.

- If $\Pr(q) = 0$, then $\Pr(p \mid q)$ is *undefined*.

- Because conditional probability is defined in terms of unconditional probability, our rules (above) for caclulating unconditional probabilities will allow us to calculate all conditional probabilities as well.

- Indeed, we can not only calculate (numerical) conditional and unconditional probabilities, using (numerical) STTs, we can also *prove (all) general facts* about conditional and unconditional priobabilities *via* (generic) STTs, and (general) facts about high-school algebra.

- Here are some *general* claims to prove *via* our definitions + HS algebra.

  1. $\Pr(X \vee Y) = \Pr(X) + \Pr(Y) - \Pr(X \,\&\, Y)$.
  2. $\Pr(X \to Y) \geq \Pr(Y \mid X)$.
  3. If $\Pr(X \mid Y) = \Pr(X)$, then $\Pr(Y \mid X) = \Pr(Y \mid {\sim}X)$.

Inductive Strength I

- *Deductively valid* arguments have the following feature:
  - If the premises of a *valid* argument are all true, then this *guarantees* that the conclusion of the argument is also true.
- *Inductively strong* arguments do *not* have this feature. But,
  - If the premises of a *strong* argument are all true, then this *makes it probable* that the conclusion of the argument is also true.
- Two other ways to express validity:
  - $P \therefore C$ is *valid* iff $P \to C$ is *necessary* (*i.e.*, *logically true*).
  - $P \therefore C$ is *valid* iff $P \,\&\, {\sim}C$ is *impossible* (*i.e.*, *logically false*).
- This suggests a natural idea for how to explicate "$P \therefore C$ is *strong*."
- **Skyrms's First Proposal.** $P \therefore C$ is *strong* iff $P \to C$ is *probable*. Or, equivalently, $P \therefore C$ is *strong* iff $P \,\&\, {\sim}C$ is *improbable*.

Inductive Strength II

- This first proposal of Skyrms will not do, because $P \to C$ can be probable (*i.e.*, $P \,\&\, {\sim}C$ can be *improbable*) *even if there is no relation of inductive support between P and C*. Consider the following case:
  - $P \overset{\text{def}}{=}$ "There is a 2000-year-old man in Cleveland,"
  - $C \overset{\text{def}}{=}$ "There is a *3-headed* 2000-year-old man in Cleveland."
- In this case, P *alone* is highly *im*probable. That is, ${\sim}P$ *alone* is highly *probable*. For this reason, ${\sim}P \vee C$ (*i.e.*, $P \to C$) is highly probable.

\+ But, this is true *despite the fact that there is not a strong relation of support between P and C*. *That's* what "strength" is meant to capture.

- This leads Skyrms (and most others in this literature) to adopt the following alternative explication of "inductive strength".
- **Skyrms's Second Proposal**. $P \therefore C$ is *strong* iff C is probable *given that* (*i.e.*, *on the supposition that*) P is true.

Inductive Strength III

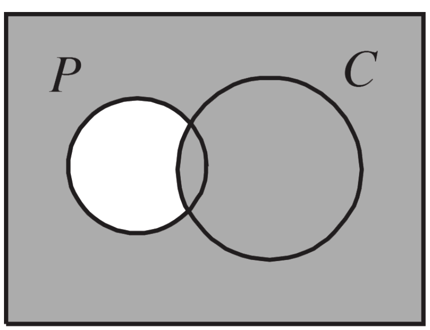
- Skyrms considers the following proposal for "inductive strength":

  Proposal #1. An argument $P \therefore C$ is inductively strong just in case the claim $P \rightarrow C$ is *probable*.

- This first proposal is inadequate, since an argument will be judged as strong if P is improbable (or C is probable). He moves to the following:

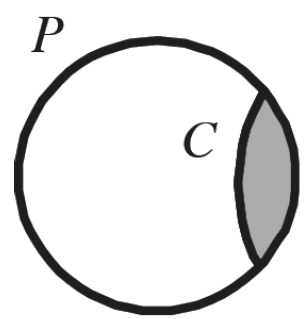
  Proposal #2. An argument $P \therefore C$ is inductively strong just in case C is probable, *given that* (*i.e.*, *on the supposition that*) P is true.

- It helps to visualize examples in which these two proposals *diverge*.



Proposal #1 (strong)

vs

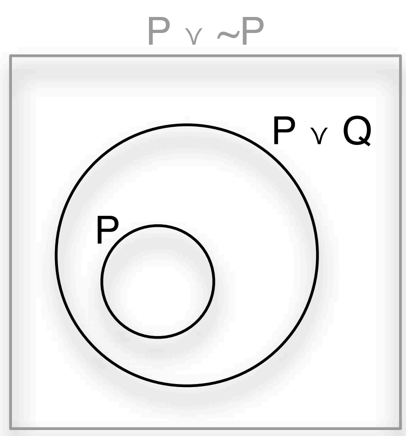


Proposal #2 (weak)

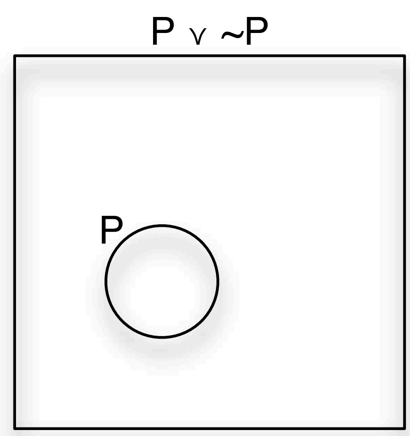
Inductive Strength IV

$$(1) \quad \begin{array}{c} P \vee Q. \\ \therefore P \end{array} \qquad\qquad (2) \quad \begin{array}{c} P \vee \sim P. \\ \therefore P \end{array}$$

- We can picture these two arguments, as follows.



(1)



(2)

- If we apply proposal #2 to (1) and (2), we get the intuitively correct verdict that (1) *is stronger than* (2).

☞ The proportion of P-worlds among the $P \vee Q$ worlds is *greater than* the proportion of P worlds among the $P \vee \sim P$ worlds.

Inductive Strength V

- The probability of C ***given that*** P is a much better guide to the inductive strength of "$P \therefore C$" than the probability of $P \rightarrow C$.
- But, there is still something lacking in Skyrms's second proposal.
- This defect can be illustrated *via* the following inductive argument.

  (P) Fred Fox (who is a man) is on birth control pills.

  Therefore, (C) Fred Fox (who is a man) will not get pregnant.

- The probability of C *given that* P is very high (as is the probability that $P \rightarrow C$). So, proposal #2 (and proposal #1) says "$P \therefore C$" is *strong*.
- But, intuitively, P is *irrelevant* to C, and so (intuitively) P *does not provide evidence in favor of* C. This suggests a third proposal.

  Proposal #3. "$P \therefore C$" is strong just in case (1) the probability of C *given that* P is *high*, *and* (2) P is *positively relevant to* C — *i.e.*, the probability of C *given that* P is high***er*** than the probability of C.

Inductive Strength VI

- The third proposal adds a *positive relevance* requirement.
- It is helpful to think about examples involving *games of chance*. Suppose card c is going to be sampled from a standard deck of cards.
- The probability that c is a spade (C), *given that* c is black (P) is $\frac{1}{2}$. I will abbreviate this *conditional probability* claim as: $\Pr(C \mid P) = \frac{1}{2}$.
- This is *not high* (*i.e.*, it is *not greater than* $\frac{1}{2}$). But, it *is* high***er*** than the probability that c is a spade (*i.e.*, the probability of C), which is $\frac{1}{4}$.
- So, in this case, P *is positively relevant to* C (note: the probability of C *given that* $P = \frac{1}{2}$, which is greater than the probability of $C = \frac{1}{4}$).
- So, "$P \therefore C$" does *not* come out *strong* on proposal #3, since $\Pr(C \mid P)$ is *not* high. But it *does* satisfy the *positive relevance* requirement. That is:
  - $\Pr(C \mid P) = \frac{1}{2}$, which is *not* high.
  - But, $\Pr(C \mid P) > \Pr(C) = \frac{1}{4}$, so P *is positively relevant to* C.

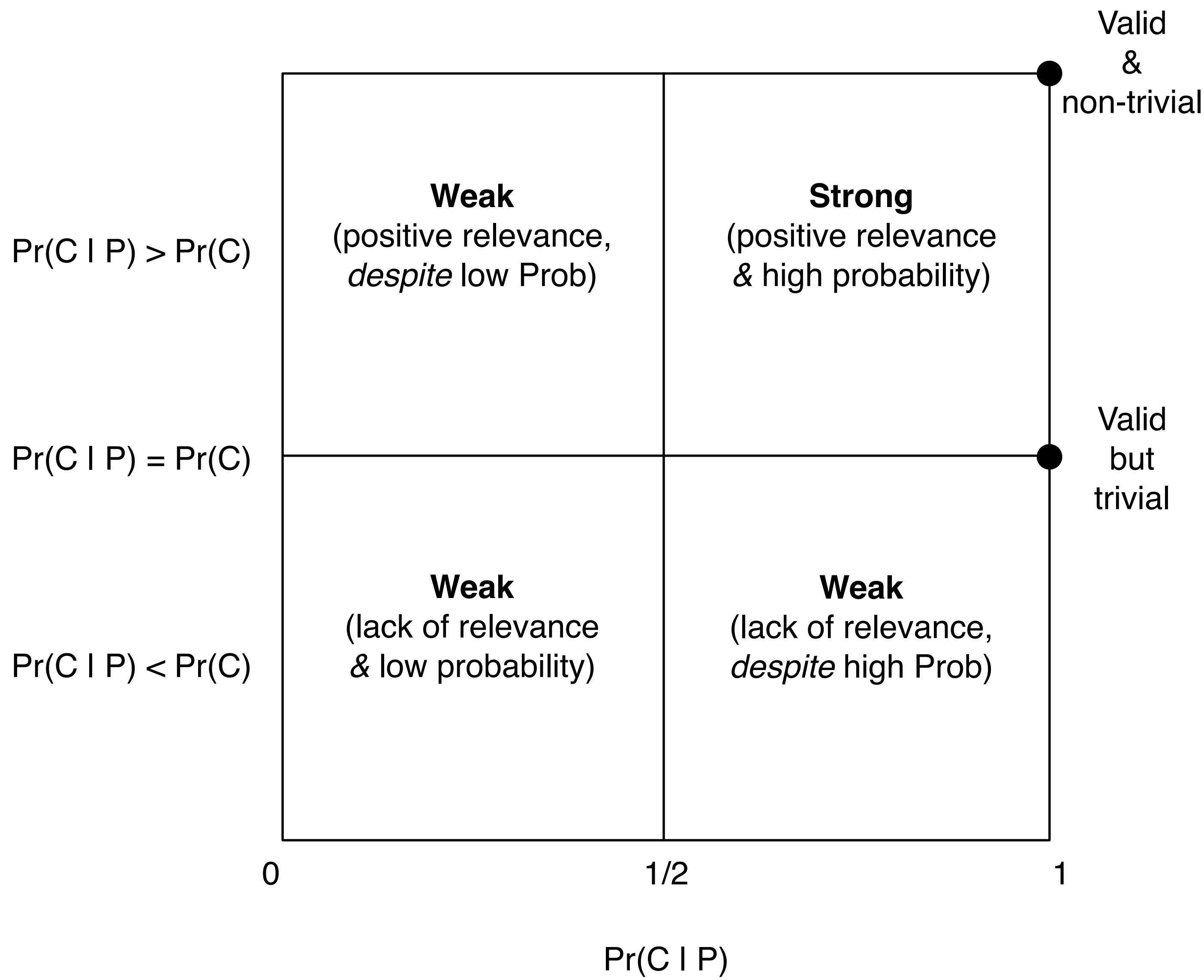
Inductive Strength VII

- With our $\Pr(\cdot)$ notations in hand, we can restate proposal #3.
- **Proposal #3.** An argument $P \therefore C$ is *inductively strong* iff

  (1) C is probable, *given* P, *i.e.*, $\Pr(C \mid P) > \frac{1}{2}$, *and*

  (2) P is *positively relevant* to C, *i.e.*, $\Pr(C \mid P) > \Pr(C)$.
- This proposal is superior to Skyrms's, as it requires *both* that $\Pr(C \mid P)$ be *high* ($> \frac{1}{2}$), *and* that $\Pr(C \mid P)$ be high*er* than $\Pr(C)$. This means P has to *raise the probability of C to a number that is greater than* $\frac{1}{2}$.
- I won't offer a measure of *degree* of inductive strength [$\mathfrak{c}(C, P)$], but, presumably, $\mathfrak{c}(C, P)$ would be some function of $\Pr(C \mid P)$ and $\Pr(C)$.
- The important thing is that one must think about *two factors* when assessing whether an argument $\ulcorner P \therefore C \urcorner$ is strong.
  - **Factor #1.** $\Pr(C \mid P)$ must be *high* ($> \frac{1}{2}$).
  - **Factor #2.** $\Pr(C \mid P)$ must be high***er*** than $\Pr(C)$.

Our “Two-Factor” Approach to Inductive Strength: A Chart



Inductive Strength and Trivial *vs* Non-Trivial Validities

- Usually, valid arguments will come out strong on our proposal. This is because most valid arguments are non-trivial, and so they satisfy both of our requirements of high probability and positive relevance.
- But, there are some (fringe) validities that are *trivial*, and *not* strong.
- Recall that the following two LSL arguments are *valid*:

  (i) $P \,\&\, {\sim}P \therefore Q$.

  (ii) $P \therefore Q \vee {\sim}Q$.

- I will call these *trivial* validities, because they do *not* count as strong on our *two-factor* approach. Argument (ii) is easier to think about.
- What is $\Pr(Q \vee {\sim}Q \mid P)$? Because $Q \vee {\sim}Q$ is a *tautology*, we will have $\Pr(Q \vee {\sim}Q \mid P) = 1$. So, on Proposal #2, argument (ii) comes out strong. But, this is incorrect, since $\Pr(Q \vee {\sim}Q \mid P) = \Pr(Q \vee {\sim}Q) = 1$.
- That is, P is *irrelevant to* $Q \vee {\sim}Q$. So, (ii) is *not* a strong argument.