The Story of Inductive Logic

Branden Fitelson

1 Deductive Logic and its Limitations

We will be using **classical sentential (*viz.*, truth-functional/Boolean) logic** as our background, deductive logical theory. This theory (*viz.*, the truth-table method we will be using to reason, semantically, about it) traces back to Peirce [8] (and, later, Wittgenstein [17]). The basic units of analysis in sentential logic are **atomic sentences**. These are meant to be declarative sentences which contain no (sentential) logical connectives. We will use capital letters: A, B, C, ... to denote atomic sentences. The only other elements of the language of sentential logic (LSL) are the (sentential) logical connectives (hereafter, **the connectives**) themselves. The meanings of the connectives are given by the following **truth-table definitions**.

p	$\sim p$
T	F
F	T

p	q	$p \,\&\, q$
T	T	T
T	F	F
F	T	F
F	F	F

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

p	q	$p \to q$
T	T	T
T	F	F
F	T	T
F	F	T

p	q	$p \equiv q$
T	T	T
T	F	F
F	T	F
F	F	T

The meaning of a complex LSL sentence is then given by its truth-table, which can be constructed from these basic definitions. For instance, consider the sentence $X \,\&\, (Y \equiv Z)$. Its truth-table is as follows:

X	Y	Z	$X \,\&\, (Y \equiv Z)$
T	T	T	T
T	T	F	F
T	F	T	F
T	F	F	T
F	T	T	F
F	T	F	F
F	F	T	F
F	F	F	F

We will think of the rows of a truth-table as representing **possible worlds**. An LSL sentence p is said to be a **tautology** if is true in every possible world — *i.e.*, if every row of p's truth-table is T; p is a **contradiction** if every row of p's truth-table is F; otherwise, p is **contingent**.

An (LSL) **argument** is a set of LSL sentences, one of which is the **conclusion** (the remainder are **premises**). We will simplify things by treating every argument as if it has a single premise $\mathcal{P}$ (just take the conjunction of the individual premises) and conclusion C. An argument $\mathcal{P} \therefore C$ is **valid** iff the conjunction $\mathcal{P} \,\&\, \sim C$ is a contradiction. [Equivalently, $\mathcal{P} \therefore C$ is valid iff the conditional $\mathcal{P} \to C$ is a tautology.] We will use the notation $\mathcal{P} \vDash C$ to express the claim that the argument $\mathcal{P} \therefore C$ is valid.

I will focus on two kinds of **limitations** (and/or shortcomings) of (sentential) deductive logic.

- **Semantic Limitations** (shortcomings of the truth-functional definitions of the connectives)

  – The logical constants of LSL are **truth-functional**. As a result, they only capture *some* of the meaning of various natural language connectives. For instance, in English, we have various *non-truth-functional* sentential connectives, *e.g.*, ⌜*p but q*⌝, ⌜*p because q*⌝; and, ⌜*if p, then q*⌝ (the English indicative conditional). We symbolize ⌜*p but q*⌝ and ⌜*p because q*⌝ as ⌜$p \,\&\, q$⌝. But, this only captures *some* of their meaning. And, we symbolize ⌜*if p, then q*⌝ as ⌜$p \to q$⌝. But, this only captures *some* of the meaning of the English conditional.[1] I'll explain in a later lecture how probability helps us to model some of the non-truth-functional meaning of the English connectives.

- **Logical Limitations** (shortcomings of the truth-functional definition of validity)

  – **Over-Generation Cases** (the so-called "**Paradoxes of Entailment**"). According to the our (truth-functional) definition of validity, the following two forms of argument are generally valid.

  $$P \therefore Q \vee {\sim}Q \qquad\qquad P \,\&\, {\sim}P \therefore Q$$

  This is easily seen by examining the following truth-tables:

P	Q	$P \,\&\, {\sim}P$	$Q \vee {\sim}Q$
T	T	F	T
T	F	F	T
F	T	F	T
F	F	F	T

  There is no possible world in which 'P' is true *and* '$Q \vee {\sim}Q$' is false (because there is no possible world in which '$Q \vee {\sim}Q$' is false); and, there is no possible world in which '$P \,\&\, {\sim}P$' is true *and* 'Q' is false (because there is no possible world in which '$P \,\&\, {\sim}P$' is true). Intuitively, there seems something odd about these arguments. For instance, P (the moon is made of green cheese) could be *intuitively irrelevant* to $Q \vee {\sim}Q$ (either there is a Santa Claus or there is not).

  Some logicians have proposed criteria of "deductive relevance" (*e.g.*, the "variable-sharing" criterion [12]) to try to restrict the classical definition of validity — so as to rule out cases like these. I don't think this is necessary. Indeed, as I will argue below, I think (probabilistic) inductive logic provides a better explanation of the shortcomings of these "paradoxes of entailment."

  – **Under-Generation Cases** (the **Bachelor Argument**). Let $B \overset{\text{def}}{=}$ "John in a bachelor;" and, let $M \overset{\text{def}}{=}$ "John is married." We can symbolize the argument "John is a bachelor. Therefore, John is unmarried." as $B \therefore {\sim}M$. This argument is invalid, because there is a *logically* possible world in which B and M are both assigned the truth-value True. But, once we understand the *meanings* of the propositions expressed by B and M, we realize that such worlds are *conceptually impossible*.[2] Moreover, in this case, adding a "relevance criterion" will only *reinforce* the verdict of "invalidity." After all, from a *variable-sharing* perspective, B *is* "irrelevant" to ${\sim}M$. Later on, I will argue that (probabilistic) inductive logic gives us a simple way to explain why this argument is strong.

[1] There are 16 binary truth-functions ($\star$). They are as follows [8].

p	q	$\top$	NAND	$\to$	${\sim}p$	FI ($\leftarrow$)	${\sim}q$	$\equiv$	NOR	$\vee$	NIFF	q	NFI	p	NIF	$\&$	$\bot$
T	T	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F
T	F	T	T	F	F	T	T	F	F	T	T	F	F	T	T	F	F
F	T	T	T	T	T	F	F	F	F	T	T	T	T	F	F	F	F
F	F	T	T	T	T	T	T	T	T	F	F	F	F	F	F	F	F

Exercise$_1$: show that *the only one* of these 16 truth-functions ($\star$) that satisfies all three of the following desiderata is '$\to$' (depicted in red). (i) *modus ponens* is valid $[p \,\&\, (p \star q) \vDash q]$, (ii) affirming the consequent is invalid $[q \,\&\, (p \star q) \nvDash p]$, (iii) $p \star p$ is a tautology. See Lecture #11 in my (YouTube) Logic lecture series (second barcode at the top of this handout) for a more complete explanation.

[2] Perhaps "bachelor" can be conceptually analyzed as a conjunction of concepts, including "unmarried." In that case, sentential logic could deliver a valid verdict on a finer-grained analysis of the argument. Perhaps, but this strategy won't work for all cases. *E.g.*, "John knows that it's raining. ∴ It's raining." It seems that no suitable conceptual analysis of "knowledge" is forthcoming [16].

2 Probability Calculus: An Algebraic Approach

The **probability calculus** can be thought of as a numerical extension/generalization of sentential logic. We will add to our truth-tables another column, which represents a **probability function** $\Pr(\cdot)$ over the possible worlds represented by the rows of the truth-table. We call such extended truth-tables **probability tables** (PTs) [3, 13]. A probability function $\Pr(\cdot)$ is simply an assignment of real numbers a_i to each possible world w_i. To be more precise, we will associate with each possible world w_i a **state description** s_i, which is just an LSL conjunction that expresses exactly which atomic sentences are true (and which are false) in world w_i. Then, our probability function $\Pr(\cdot)$ will assign numerical values a_i to each state description s_i. Here is a (generic) probability table for the three-atom (X, Y, Z) case.

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

The only constraints on the **basic probabilities** a_i are the following two:

(1) Each of the a_i must lie on the unit interval $[0, 1]$.

$$a_1, \ldots, a_{2^n} \in [0, 1]$$

(2) The a_i must sum to 1.

$$\sum_{i=1}^{2^n} a_i = 1$$

With these definitions in place, we can now define the (unconditional) probability $\Pr(p)$ of an arbitrary LSL sentence p, in terms of the basic probabilities a_i, as follows.

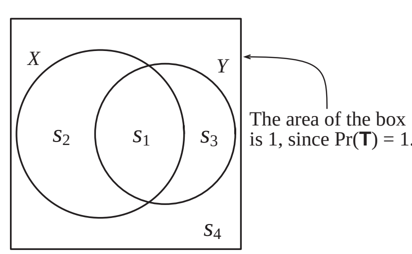
$$\Pr(p) \overset{\text{def}}{=} \sum_{s_i \vDash p} \Pr(s_i) = \sum_{s_i \vDash p} a_i$$

In words: the probability of p is the sum of the basic probabilities of the state descriptions s_i which entail p.[3] Applying this definition to our example $p \overset{\text{def}}{=} X \,\&\, (Y \equiv Z)$ yields:

$$\Pr(X \,\&\, (Y \equiv Z)) \overset{\text{def}}{=} \sum_{s_i \vDash X \& (Y \equiv Z)} a_i = a_1 + a_4$$

We can also use **probabilistic venn diagrams** (PVDs) to visualize (specific, numerical) probability functions. Here is a concrete, numerical example of an PT and its associated PVD.

X	Y	States	$\Pr(s_i)$
T	T	s_1	$a_1 = {}^4\!/_{24} \approx 0.166$
T	F	s_2	$a_2 = {}^6\!/_{24} = 0.25$
F	T	s_3	$a_3 = {}^3\!/_{24} = 0.125$
F	F	s_4	$a_4 = {}^{11}\!/_{24} \approx 0.458$

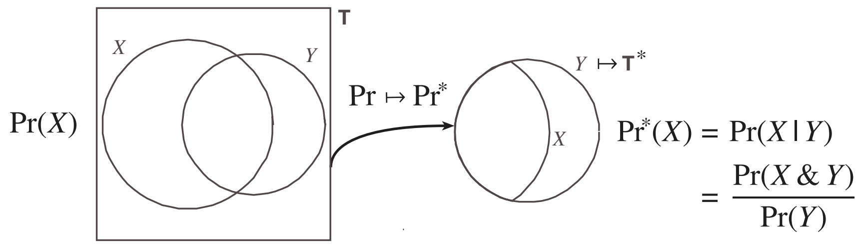


[3]If p is a contradiction, then $\Pr(p) = 0$, because there are *no* state descriptions which entail a contradiction. Note: *the converse of this claim is false.* It is consistent with the rules of probability calculus that a logically contingent claim — *e.g.*, "John is a married bachelor" — be assigned probability *zero*. As we will see when we get to applications, this is a *feature* — not a bug! In infinite algebras, even contingent *empirical* claims can have probability zero. Such cases/models are beyond the scope of this story [9, 7].

The only thing left to define is the **conditional probability** $\Pr(p \mid q)$ of one LSL sentence (p), *given* another (q). This is defined using the following **Ratio Formula** [7].

$$\Pr(p \mid q) \overset{\text{def}}{=} \frac{\Pr(p \,\&\, q)}{\Pr(q)}, \text{ provided that } \Pr(q) > 0.$$

The rationale for defining conditional probability using the ratio formula can be visually explained using PVDs. 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 *indicatively 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") PVD*. This is like *moving to a new* $\Pr^*$*-function, according to which* $\Pr^*(Y) = 1$. And, this process naturally yields the ratio formula.



Here's an example to illustrate the definition of conditional probability (using our 3-atom PT, above).

$$\Pr(X \equiv Y \mid X \equiv Z) = \frac{\Pr((X \equiv Y) \,\&\, (X \equiv Z))}{\Pr(X \equiv Z)} = \frac{\Pr((X \,\&\, Y \,\&\, Z) \vee (\sim X \,\&\, \sim Y \,\&\, \sim Z))}{\Pr(X \equiv Z)} = \frac{a_1 + a_8}{a_1 + a_3 + a_6 + a_8}$$

With these definitions in place, we are now in a position to (a) prove theorems of probability calculus, and (b) produce counterexamples to claims that are not theorems of probability calculus. We will call this **the algebraic method** for reasoning about probability calculus. The algebraic method involves two steps.

- **First Step**: **Algebraic Translation**. Translate probabilistic expressions or claims into their algebraic equivalents (using our definitions of unconditional and conditional probability).
- **Second Step**: **Algebraic Calculation**. Here, we **either** "*plug & chug*" a given *numerical* assignment to the state probabilities a_i **or** we *reason generally/abstractly* about the a_i, using constraints (1) and (2). Iff the algebraic translation of a probabilistic statement α is a *theorem of algebra* — for any assignment to the a_i satisfying (1) and (2) — then α is a **theorem of probability calculus** [3].[4]

Exercise$_2$. Consider the following numerical probability function.

X	Y	States	$\Pr(s_i)$
T	T	s_1	$a_1 = 7/16$
T	F	s_2	$a_2 = 1/16$
F	T	s_3	$a_3 = 5/16$
F	F	s_4	$a_4 = 3/16$

Use this numerical probability function to show that none of (3)–(5) are theorems of probability calculus. That is, show that the this numerical probability function is a case in which claims (3)–(5) all *fail* to hold.

(3) $\Pr(X \mid Y) = \Pr(X \mid \sim Y)$.

(4) $\Pr(X \equiv Y) \geq \Pr(Y \mid X)$.

(5) $\Pr(X \vee Y) = \Pr(X) + \Pr(Y)$.

[4]Because probability calculus can be expressed in the theory of real closed fields, it is *decidable* [3]. I have written a *Mathematica* package (PrSAT) implementing such a decision procedure. It can be downloaded from `http://fitelson.org/PrSAT/`. Thanks to my student Koissi Adjorlolo, we now have a web-based (alpha) version: `https://imapersonman.github.io/PrSAT/`.

Exercise$_3$. Prove **Bayes's Theorem** (*viz.*, the following equation), using the def. of conditional probability.

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \cdot \Pr(H)}{\Pr(E \mid H) \cdot \Pr(H) + \Pr(E \mid {\sim}H) \cdot \Pr({\sim}H)}$$

Terminological notes on Bayes's Theorem: $\Pr(H \mid E)$ is called the *posterior* of the hypothesis H (given evidence E), $\Pr(E \mid H)$ is called the *likelihood* of H (relative to E), $\Pr(H)$ is called the *prior* (or *base rate*) of the hypothesis H, and $\Pr(E \mid {\sim}H)$ is called the *catch-all likelihood* (of ${\sim}H$, relative to E). Sometimes, the likelihood is called the *true positive rate*, and the catch-all likelihood is called the *false positive rate*. So, Bayes's Theorem tells us how the posterior is determined by the prior, the likelihood, and the catch-all likelihood (or the prior, the true positive rate and the false positive rate). We will use Bayes's Theorem in an application.

Exercise$_4$. Prove (algebraically) that the following is a theorem of probability calculus

$$\Pr(X \to Y) \geq \Pr(Y \mid X).$$

3 A 2-Dimensional (Probabilistic) Theory of Argument Strength: Part I

Inductive logic [4] aims to generalize the consequence relation ($\vDash$) of deductive logic, using the probability calculus. Ideally, we would like to define a function $s(C, \mathcal{P})$ — using a probability function $\Pr(\cdot)$ defined over the atomic sentences contained in the argument — which measures (on a suitable numerical scale) the **inductive strength** of the argument $\mathcal{P} \therefore C$. If $s(C, \mathcal{P})$ is going to quantitatively generalize $\mathcal{P} \vDash C$, then we would — at the very least — want it to satisfy the following two constraints.

- **(Non-Paradoxical) Entailments (Max).** If $\mathcal{P}$ (non-paradoxically) entails C, then $s(C, \mathcal{P})$ is *maximal*.
- **(Non-Paradoxical) Refutations (Min).** If $\mathcal{P}$ (non-paradoxically) refutes C, then $s(C, \mathcal{P})$ is *minimal*.

Brian Skyrms [11] considers two specific proposals for $s(C, \mathcal{P})$.

Skyrms's First Proposal. Recall that an argument $\mathcal{P} \therefore C$ valid iff the conditional $\mathcal{P} \to C$ is a tautology. Another way to put this is to say that $\mathcal{P} \therefore C$ valid iff $\mathcal{P} \to C$ is *necessary*. A simple way to generalize this, probabilistically, is to say that $\mathcal{P} \therefore C$ is **strong** iff $\mathcal{P} \to C$ is *probable*. This leads to the following proposal.

$$s_1(C, \mathcal{P}) \stackrel{\text{def}}{=} \Pr(\mathcal{P} \to C)$$

Unfortunately, this proposal is inadequate, because it violates **(Non-Paradoxical) Refutations (Min)**. According to this proposal, some (non-paradoxical) refutations will be deemed *strong*. To see why, consider: $X \therefore {\sim}X$. Intuitively, this should not be a strong argument. But, according to the first proposal, we have

$$s_1({\sim}X, X) \stackrel{\text{def}}{=} \Pr(X \to {\sim}X) = \Pr({\sim}X \vee {\sim}X) = \Pr({\sim}X) = 1 - \Pr(X).$$

So, if X is improbable, then $s_1({\sim}X, X)$ is high, and proposal #1 will judge $X \therefore {\sim}X$ to be a strong argument.

Skyrms's Second Proposal. Skyrms rightly rejects the first proposal, and opts for the following alternative.

$$s_2(C, \mathcal{P}) \stackrel{\text{def}}{=} \Pr(C \mid \mathcal{P})$$

This second proposal satisfies both **(Non-Paradoxical) Entailments (Max)** and **(Non-Paradoxical) Refutations (Min)** (why?). So, in this sense, it is clearly superior to the first proposal. However, there is still something lacking in this account. The missing element is revealed by the following 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.

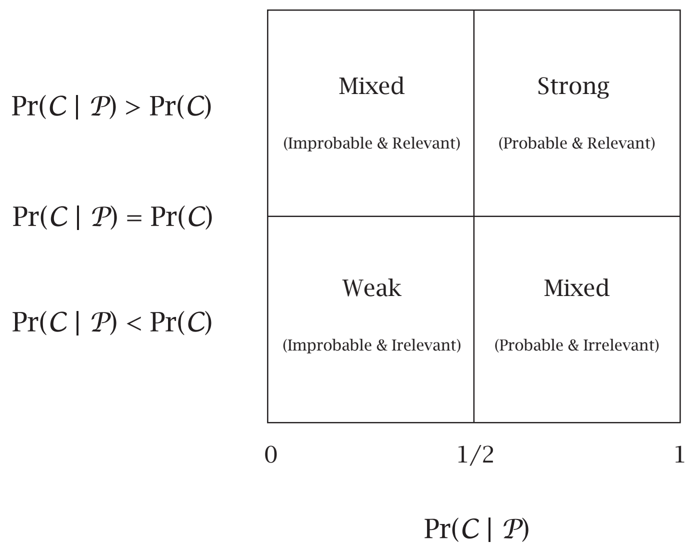
Intuitively, $\Pr(C \mid P)$ is high in this case. Therefore, according to Skyrms's second proposal, this is judged to be a strong argument. This seems wrong. While it is true that the conditional probability $\Pr(C \mid P)$ is high, the premise is (intuitively) *irrelevant* to the conclusion. That is, intuitively, we have $\Pr(C \mid P) = \Pr(C)$. As a result, this example reveals that Skyrms's second proposal violates the following plausible constraint.

- **Strong Arguments are Positively Relevant**. If $\mathcal{P} \therefore C$ is strong, then $\Pr(C \mid \mathcal{P}) > \Pr(C)$.

What all of this reveals is that ***there are two dimensions of argument strength***: *probability* & *relevance*. That suggests the following proposal for a qualitative conception of what it means to be a *strong* argument.

> **Strength (Qualitative Defintion)**. $\mathcal{P} \therefore C$ is strong iff *both* (1) the probability of C *given that* $\mathcal{P}$ is *high*, *and* (2) $\mathcal{P}$ is *positively relevant to* C. Or, more formally, if both of the following conditions are met.
>
> – $\Pr(C \mid \mathcal{P})$ is *high* (*i.e.*, $\Pr(C \mid \mathcal{P}) > t$, for some $t \geq 1/2$).
> – $\Pr(C \mid \mathcal{P})$ is high*er* than $\Pr(C)$ (*i.e.*, $\Pr(C \mid \mathcal{P}) > \Pr(C)$).

We can visualize the space of argument strengths with the following 2-D (four quadrant) diagram. The x-axis is $\Pr(C \mid \mathcal{P})$. The y-axis has three categories: *positive* relevance, *ir*relevance, and *negative* relevance.



4 A 2-Dimensional (Probabilistic) Theory of Argument Strength: Part II

So far, we only have a *qualitative* (three-category) relevance dimension. It would be nice to have a *numerical probabilistic relevance measure* $\mathfrak{R}(C, \mathcal{P})$. As it turns out, there are *many* such measures. Here are a few examples from the literature (see [5] for a comprehensive survey). Note: they are all on a $[-1, 1]$ scale.

- The *Difference*: $d(C, \mathcal{P}) \overset{\text{def}}{=} \Pr(C \mid \mathcal{P}) - \Pr(C)$
- The *Ratio*: $r(C, \mathcal{P}) \overset{\text{def}}{=} \dfrac{\Pr(C \mid \mathcal{P}) - \Pr(C)}{\Pr(C \mid \mathcal{P}) + \Pr(C)}$
- The *Likelihood-Ratio*: $l(C, \mathcal{P}) \overset{\text{def}}{=} \dfrac{\Pr(\mathcal{P} \mid C) - \Pr(\mathcal{P} \mid {\sim}C)}{\Pr(\mathcal{P} \mid C) + \Pr(\mathcal{P} \mid {\sim}C)}$
- The *Normalized-Difference*: $s(C, \mathcal{P}) \overset{\text{def}}{=} \Pr(C \mid \mathcal{P}) - \Pr(C \mid {\sim}\mathcal{P}) = \dfrac{1}{\Pr({\sim}\mathcal{P})} \cdot d(C, \mathcal{P})$

Interestingly, *none of these measures agrees with any of the others on comparative judgments of relevance* (*i.e.*, no pair is *comparatively equivalent*). That is to say, for every pair of relevance measures $\langle \mathcal{R}_1, \mathcal{R}_2 \rangle$ from the above list of four measures, there will exist pairs of arguments $\langle \mathcal{P}_1, C_1 \rangle$ and $\langle \mathcal{P}_2, C_2 \rangle$ such that

$$\mathcal{R}_1(C_1, \mathcal{P}_1) \geq \mathcal{R}_1(C_2, \mathcal{P}_2); \textbf{\textit{but}}, \mathcal{R}_2(C_1, \mathcal{P}_1) < \mathcal{R}_2(C_2, \mathcal{P}_2)$$

Exercise$_5$: Consider the following numerical probability distribution over the two atoms X and Y.

X	Y	$\Pr(s_i)$
T	T	$a_1 = {}^{33}/_{64}$
T	F	$a_2 = {}^{15}/_{64}$
F	T	$a_3 = {}^{3}/_{64}$
F	F	$a_4 = {}^{13}/_{64}$

Show that the following three claims are true, according to this probability distribution.

(6) $d(X, Y) > d(Y, X)$.

(7) $s(X, Y) < s(Y, X)$.

(8) $r(X, Y) = r(Y, X)$.

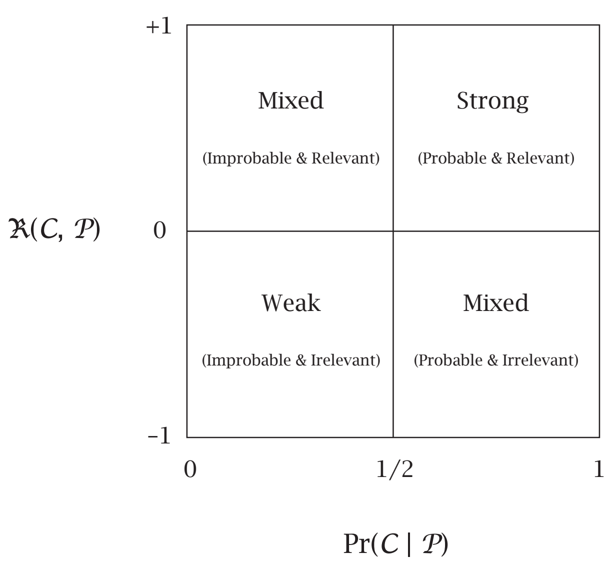
This establishes that d, s, and r are *not comparatively equivalent*. Similar examples can be given to show that l is not equivalent to any of the other three measures. But, there is an easier way to see why l must disagree with the other three measures. For l is the only measure satisfying the following *desideratum* (**Exercise**$_6$) — that $\mathfrak{R}$ be a *numerical generalization of (non-paradoxical) entailment and refutation* [4, 6].

$$\mathfrak{R}(C, \mathcal{P}) \text{ should be } \begin{cases} +1 & \text{if } \mathcal{P} \vDash C \text{ (non-paradoxically).} \\ > 0 \text{ (positive relevance)} & \text{if } \Pr(C \mid \mathcal{P}) > \Pr(C). \\ = 0 \text{ (irrelevance)} & \text{if } \Pr(C \mid \mathcal{P}) = \Pr(C). \\ < 0 \text{ (negative relevance)} & \text{if } \Pr(C \mid \mathcal{P}) < \Pr(C). \\ -1 & \text{if } \mathcal{P} \vDash {\sim}C \text{ (non-paradoxically).} \end{cases}$$

Remember that we want our (logical) assessments of argument strength to involve *quantitative generalizations of entailment* (and refutation). This is why we want (non-paradoxical) entailments to receive maximal relevance scores and (non-paradoxical) refutations to receive minimal relevance scores (*viz.*, extremal *negative* relevance scores). Since l is the only measure satisfying our logical desideratum, we will adopt it as our measure of degree of relevance of an argument. To wit, we will adopt the following definition.

$$\mathfrak{R}(C, \mathcal{P}) \overset{\text{def}}{=} \frac{\Pr(\mathcal{P} \mid C) - \Pr(\mathcal{P} \mid {\sim}C)}{\Pr(\mathcal{P} \mid C) + \Pr(\mathcal{P} \mid {\sim}C)}$$

With this relevance measure in hand, we now have the following fully quantitative 2-D map of strengths.



This allows us to map the strength of any argument — relative to any particular, evaluative probability distribution — as an ordered pair of real numbers (*i.e.*, as a *point* in our 2-D strength map).[5]

Exercise$_7$. Map the simple arguments $X \therefore X$ and $X \therefore {\sim}X$ — assuming that X has intermediate probability.

Note that *we do not have a single numerical measure (s) of strength* — because strength has two dimensions. But, we can say some unambiguous things about some comparisons of argument strength. For instance, as we move *up and to the right* in the diagram, arguments get stronger (in both dimensions); and, as we move *down and to the left*, arguments get weaker (in both dimensions). However, if we move (*e.g.*) down and to the right in the diagram, then we have gotten *more probable but less relevant*; and, in such cases, there may not be a determinate answer to whether we have "strengthened the argument" (*all-things-considered*). Moreover, the two dimensions can come into conflict (*i.e.*, "pull us in opposite directions"); and, this can lead to some ambiguous/confusing assessments of argument strength (in "mixed" cases). In the final section, I will apply our theory to various examples from both the logic literature and the cognitive science literature.

5 Some Applications of Our 2-Dimensional Theory of Argument Strength

5.1 Applications to the Semantic Limitations of Sentential Logic

5.1.1 Probabilistic Relevance and the Meaning of 'but'

Intuitively, $\ulcorner p$ *but* $q\urcorner$ implies $\ulcorner p \,\&\, q\urcorner$. However, the converse is not true. That is, $\ulcorner p$ *but* $q\urcorner$ has non-truth-functional meaning which is over-and-above the truth-functional part of its meaning. Of course, sentential logic is powerless to capture this non-truth-functional meaning. But, at least some of this meaning can be captured in our probabilistic framework. Plausibly, $\ulcorner p$ *but* $q\urcorner$ *also* implies that *p is negatively relevant to q* — according to probability functions which encode the meaning of 'but.' That is, if $\Pr(\cdot)$ is a function which encodes the meaning of 'but', then we will (generally) have $\Pr(q \mid p) < \Pr(q)$, whenever $\ulcorner p$ *but* $q\urcorner$ is true.

5.1.2 Probabilistic Relevance and the Meaning of 'because'

Intuitively, $\ulcorner p$ *because* $q\urcorner$ implies $\ulcorner p \,\&\, q\urcorner$. After all, if p is *false* then p *does not have* an explanation; and, if q is false then *it cannot explain why* p is true. However, the converse is not true. That is, $\ulcorner p$ *because* $q\urcorner$ has non-truth-functional meaning which is over-and-above the truth-functional part of its meaning. Plausibly, $\ulcorner p$ *because* $q\urcorner$ *also* implies that *q is positively relevant to p* — according to probability functions which encode the meaning of 'because.' That is, if $\Pr(\cdot)$ is a function which encodes the meaning of 'because', then we will (generally) have $\Pr(p \mid q) > \Pr(p)$, whenever $\ulcorner p$ *because* $q\urcorner$ is true.

5.1.3 Argument Strength and the Meaning of 'if ... then ...'

Intuitively, $\ulcorner$*if* p, *then* $q\urcorner$ implies $\ulcorner p \to q\urcorner$. In order to see this, consider the contrapositive of this claim. Suppose $\ulcorner p \to q\urcorner$ is false. That is, suppose p is true and q is false. Then, intuitively, the English indicative $\ulcorner$*if* p, *then* $q\urcorner$ is also false. However, the converse is not true. That is, $\ulcorner$*if* p, *then* $q\urcorner$ has non-truth-functional meaning which is over-and-above the truth-functional part of its meaning. Plausibly, $\ulcorner$*if* p, *then* $q\urcorner$ *also* implies that *p is positively relevant to q* — according to probability functions which encode the meaning of 'if.' That is, if $\Pr(\cdot)$ is a function which encodes the meaning of 'if', then we will (generally) have $\Pr(q \mid p) > \Pr(q)$, whenever $\ulcorner$*if* p, *then* $q\urcorner$ is true. Indeed, I favor an account of $\ulcorner$*if* p, *then* $q\urcorner$ according to which $\ulcorner$*if* p, *then* $q\urcorner$ implies that *the argument $p \therefore q$ is strong.* This sort of account has been defended by various authors (a good recent example is Igor Douven [2]). On such an account (assuming our 2-D approach to argument strength), both $\Pr(q \mid p) > 1/2$ and $\Pr(q \mid p) > \Pr(q)$ should hold when $\ulcorner$*if* p, *then* $q\urcorner$ is true.

[5] Note: except for some points on the boundary, *all points in this map are possible*, for *some* evaluative probability distribution(s).

5.2 Applications to the Logical Limitations of Sentential Logic

5.2.1 The First "Paradox of Entailment"

The first "paradox of entailment" is

$$P \therefore Q \vee {\sim}Q.$$

In order to evaluate this argument, we need to determine $\Pr(Q \vee {\sim}Q \mid P)$ and $\Pr(Q \vee {\sim}Q)$. These will both be equal to 1 — provided only that $\Pr(P) > 0$. So, this argument will (almost always) have *maximal probability*; but, it will (almost always) be *irrelevant*. [Where will it be located on our 2-D map of argument strength?] As a result, this first "paradox of entailment" will (almost always) be classified as a *weak* argument.

5.2.2 The Second "Paradox of Entailment"

The second "paradox of entailment" is

$$P \,\&\, {\sim}P \therefore Q.$$

In order to evaluate the strength of this argument, we need to ask ourselves the following two questions:

(Q_1) How probable is Q — *on the indicative supposition that* $P \,\&\, {\sim}P$?

(Q_2) How probable is Q — *unconditionally*?

In general, there will be no (conceptual) difficulty in answering Q_2. But, what about Q_1? One consequence of the ratio definition of conditional probability is that $\Pr(Q \mid P \,\&\, {\sim}P)$ is *undefined*. So, according to our definitions, this argument will *not* come out strong. It will also not come out as weak, either. Both dimensions of the strength of this argument will be *undefined*. And, so, this argument will have *no* strength.[6]

5.2.3 The Bachelor Argument

Intuitively, we would like to be able to say that the argument $B \therefore {\sim}M$ is *strong*. On our account, that would mean we would need to argue that both of the following probabilistic claims are true

(9) $\Pr({\sim}M \mid B)$ is *high*,

(10) $\Pr({\sim}M \mid B)$ is high*er* than $\Pr({\sim}M)$.

Let us suppose that the probability function we are using to evaluate the strength of the argument *encodes the meanings of* "bachelor" and "married" (as it would if it reflected *our degrees of confidence*). Then, it will (generally) satisfy (9) and (10). This is because our degrees of confidence will (generally) be such that $\Pr({\sim}M \mid B) = 1$, and $\Pr({\sim}M) < 1$. So, on our account, the argument will be deemed *strong*, as desired.

5.3 Applications to Two Infamous "Reasoning Fallacies" from Cognitive Science

Kahneman & Tversky [15, 14] won the Nobel Prize in economics for their work on the psychology of probabilistic reasoning. I will apply our 2-D theory of argument strength to two of their famous case studies.

[6] To my mind, this is a reasonable verdict. For instance, does it make sense to (indicatively) suppose that *Branden weighs more than 100 pounds, and it is not the case that Branden weighs more than 100 pounds*? Many philosophers would argue that such suppositions are *defective* [18]. And, this would support the "no strength" verdict that our theory delivers.

5.4 The "Base Rate Fallacy"

> The (unconditional) probability of breast cancer is 1% for a woman at age forty who participates in routine screening. The probability of such a woman having a positive mammogram, given that she has breast cancer, is 80%. The probability of such a woman having a positive mammogram, given that she does not have breast cancer, is 10%. What is the probability that such a woman has breast cancer, given that she has had a positive mammogram in routine screening?

We can formalize this, as follows. Let $H \stackrel{\text{def}}{=}$ such a woman (age 40 who participates in routine screening) has breast cancer, and $E \stackrel{\text{def}}{=}$ such a woman has had a positive mammogram in routine screening. Then, we are given the following values for the likelihood, the catch-all likelihood, and the prior (*i.e.*, the true positive rate of the mammogram test, the false positive rate of the test, and the prior/base rate of the disease).

$$\Pr(E \mid H) = 0.8, \Pr(E \mid {\sim}H) = 0.1, \text{ and } \Pr(H) = 0.01.$$

Question: What is $\Pr(H \mid E)$? What would you guess? Most experts guess a pretty high number ($\approx$0.8). If we apply Bayes's Theorem, we get the following answer:

$$\begin{aligned}\Pr(H \mid E) &= \frac{\Pr(E \mid H) \cdot \Pr(H)}{\Pr(E \mid H) \cdot \Pr(H) + \Pr(E \mid {\sim}H) \cdot \Pr({\sim}H)} \\ &= \frac{0.8 \cdot 0.01}{0.8 \cdot 0.01 + 0.1 \cdot 0.99} \approx 0.075\end{aligned}$$

We can also use our algebraic technique to compute an answer.

E	H	$\Pr(s_i)$
T	T	$a_1 = 8/1000$
T	F	$a_2 = 99/1000$
F	T	$a_3 = 2/1000$
F	F	$891/1000$

$$\Pr(E \mid H) = \frac{\Pr(E \,\&\, H)}{\Pr(H)} = \frac{a_1}{a_1 + a_3} = 8/10$$

$$\Pr(E \mid {\sim}H) = \frac{\Pr(E \,\&\, {\sim}H)}{\Pr({\sim}H)} = \frac{a_2}{1 - (a_1 + a_3)} = 1/10$$

$$\Pr(H) = a_1 + a_3 = 1/100$$

This over-estimate of $\Pr(H \mid E)$ is usually called the "*base rate fallacy*." People tend to neglect the prior/base rate $\Pr(H)$ in their estimate of $\Pr(H \mid E)$ — *especially when E is strongly positively relevant to H.*

Exercise$_8$. Show that $\mathfrak{R}(H, E) = 7/9 \approx 0.78$, and place this argument on our map of argument strength.

Here, it is important to note that our two dimensions *pull in opposite directions*. So, when we ask *how strong the argument $E \therefore H$ is*, it is not surprising that people could be confused by this question. After all, there is a sense in which the argument is weak (it has low probability). But, there is also a sense in which the argument is strong (it has high positive relevance). I think this is relevant to explaining why people tend to over-estimate conditional probabilities in such cases.

Exercise$_9$. Show that *just the information about the prior and the catch-all likelihood already entail that the posterior is low*. Specifically, show that $\Pr(E \mid {\sim}H) = 0.1$ and $\Pr(H) = 0.01$ jointly entail that $\Pr(H \mid E) < 0.092$.

5.5 The "Conjunction Fallacy"

Here is some evidence E regarding a woman named "Linda."

(E) Linda is 31, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and she also participated in antinuclear demonstrations.

Question. Is it more probable, given E, that Linda is (B) a bank teller, or ($B \,\&\, F$) a bank teller *and* an active feminist? Formally, the question reduces to a comparison of the following two conditional probabilities:

$\Pr(B \mid E)$ *vs* $\Pr(B \,\&\, F \mid E)$. Of course, we must have $\Pr(B \mid E) \geq \Pr(B \,\&\, F \mid E)$ (why?). But, many people answer the question by saying that $\Pr(B \mid E) < \Pr(B \,\&\, F \mid E)$. So, why do people make this mistake?

Let us ask *which argument is stronger* ($E \therefore B$ or $E \therefore B \,\&\, F$). Of course, $E \therefore B$ is *more probable than* $E \therefore B \,\&\, F$. But, is it *more relevant*? Intuitively, E is *positively* (statistically) *relevant* to F, but E is *irrelevant* to B. As a result, it makes sense that E could be *more relevant to* $B \,\&\, F$ than it is to B. To be more precise, I would argue that the following two claims are plausible in the Linda case.

- **More Probable**. $\Pr(B \mid E) > \Pr(B \,\&\, F \mid E)$.
- **Less Relevant**. $\aleph(B, E) < \aleph(B \,\&\, F, E)$.

Let $\aleph(x, y \mid z)$ measure the degree to which x is relevant to y — *on the supposition that* z. More precisely, the generalized version of our likelihood ratio measure would be defined as follows.[7]

$$\aleph(x, y \mid z) \overset{\text{def}}{=} \frac{\Pr(y \mid x \,\&\, z) - \Pr(y \mid {\sim}x \,\&\, z)}{\Pr(y \mid x \,\&\, z) + \Pr(y \mid {\sim}x \,\&\, z)}.$$

It can be shown [1] that (11) and (12) jointly entail **Less Relevant**. [**Exercise**$_{10}$: Prove this (algebraically).]

(11) $\aleph(B, E) \leq 0$,

(12) $\aleph(F, E \mid B) > 0$.

Intuitively, (11) seems true, since the given evidence does *not* seem to be positively relevant to Linda's currently being a bank teller. Similarly, (12) seems right, since — (even) *supposing* Linda *is* currently a bank teller — the given evidence is (still) positively relevant to her currently being a feminist.

Once again, our two dimensions *pull in opposite directions*, which makes the comparison of argument strengths in this case ambiguous (and confusing). Generally, when these conflicts occur, people will often revert to the relevance assessment (rather than the probability assessment) of the argument. Moreover, this deference to relevance (over probability) often seems reasonable. In fact, information about relevance (*e.g.*, likelihood ratios or true and false positive rates) is often available, even in contexts where (prior) probabilities are not (*e.g.*, diagnostic tests, and statistical evidence in empirical science [10]).

References

[1] V. Crupi, B. Fitelson, and K. Tentori. *Probability, confirmation, and the conjunction fallacy*, 2008.

[2] I. Douven, *The Epistemology of Indicative Conditionals*, 2015.

[3] B. Fitelson, *A Decision Procedure for Probability Calculus with Applications*, 2008.

[4] ______, *Inductive Logic*, 2005.

[5] ______, *Studies in Bayesian Confirmation Theory*, 2001.

[6] J.G. Kemeny and P. Oppenheim, *Degree of factual support*, 1952.

[7] A.N. Kolmogorov, *Foundations of Probability Theory*, 1933.

[8] C.S. Peirce, *The Simplest Mathematics*, 1902.

[9] M. Rescorla, *Bayesian Defeat of Certainties*, 2025.

[10] R. Royall, *Statistical Evidence: A Likelihood Paradigm*, 1997.

[11] B. Skyrms, *Choice and Chance*, 2000.

[12] S. Standefer, *Variable-Sharing as Relevance*, 2025.

[13] M. Titelbaum, *The Fundamentals of Bayesian Epistemology*, 2023.

[14] A. Tversky and D. Kahneman, *Extensional versus intuitive reasoning: The conjunction fallacy …*, 1983.

[15] A. Tversky and D. Kahneman, *On the psychology of prediction*, 1973.

[16] T. Williamson, *Knowledge and its Limits*, 2000.

[17] L. Wittgenstein, *Tractatus Logico-Philosophicus*, 1921.

[18] S. Yalcin, *Epistemic Modals*, 2007.

[7] Our two-place relevance measure $\aleph(C, \mathcal{P})$ is a *special case* [$\aleph(C, \mathcal{P} \mid \top)$] of this three-place function — where $\top$ is a tautology.