Lectures on Inductive Logic & Confirmation

Branden Fitelson

Department of Philosophy

San José State University

branden@fitelson.org

http://fitelson.org/

Overview of Lectures

- Lecture #1: Deductive Support
  - The Classical Definition of Deductive Support and Two "Anomalies"
  - "On the supposition that" *vs* "And" in Deductive Logic
  - Toward a Quantitative Measure of Logical Support
- Lecture #2: Inductive Support as a Generalization of Deductive Support
  - Two Naive Attempts at Generalization & The Received View
  - "On the supposition that" *vs* "And" in Inductive Logic
  - Several Logical Desiderata & Our Explicatum
- Lecture #3: Measures of Inductive Support and Bayesian Confirmation
  - The Problem of Measure Sensitivity in Bayesian Confirmation Theory
  - Narrowing the Field of Bayesian Relevance Measures
  - Further Desiderata & Another Path to Our Explicatum
- Time permitting: Postscripts & Details

Deductive Support I: The Classical Definition

- According to Classical Deductive Logical Theory (for simplicity, I'll stick to propositional or sentential logic in these lectures, which is hard enough!):

  E deductively supports H iff it is *impossible* that *both*:
  - E is true, *and*
  - H is false.
- I will write "$E \vDash H$" for "E deductively supports (or *entails*) H." The *theory* of (sentential) deductive support presupposes a *Boolean algebra* $\mathcal{B}$ of (atomic) propositions ($a, b, c, \ldots$), closed under $\&, \neg, \vee, \rightarrow$ ($\top, \bot \in \mathcal{B}$ constants).
- So, $E \vDash H$ is *shorthand* for $E \vDash_{\mathcal{B}} H$, which means that — *in* $\mathcal{B}$ — $E \,\&\, \neg H$ *has the same truth table as a contradiction* $\bot = p \,\&\, \neg p$ (some atomic $p \in B$).
- This relativity of $\vDash$ to $\mathcal{B}$ is important. Consider E = "John is a bachelor" and H = "John is unmarried". If, in $\mathcal{B}$, we can express E as '$m \,\&\, u$' and H as 'u', then $E \vDash_{\mathcal{B}} H$. But, if $\mathcal{B}$ is more "coarse grained" ($E = e$, $H = h$), $E \nvDash_{\mathcal{B}} H$.

Deductive Support II: Two "Anomalous" Instances

- On the classical definition of deductive support, we have ($\forall X, Y \in \mathcal{B}$):

(1) $X \,\&\, \neg X$ deductively supports Y (since $\bot \vDash Y$)

(2) X deductively supports $Y \vee \neg Y$ (since $X \vDash \top$)

- These two "anomalous" cases of deductive support have generated a great deal of controversy. Case (1) spawned the (now vast) field of *Relevance Logic* [1], and case (2) has inspired various forms of *Intuitionism* [10].
- These controversies are quite understandable. In what sense does a contradiction *support* an arbitrary proposition? And, in what sense does an arbitrary proposition *support* a tautology? We'll return to these later.
- I will not discuss the non-Boolean accounts of deductive support that have been presented in the literature. I will argue that there is *no need* to move to a "non-Boolean logic" to avoid these anomalies in the theory of *support*.

Deductive Support III: Moving Toward "Quantitative" Support

- Here's a measure of the "degree to which E deductively supports H (*in* $\mathcal{B}$)":

  Definition. Let $\delta_{\mathcal{B}}(H, E)$ be the degree to which E deductively supports H (*in* $\mathcal{B}$). We define $\delta_{\mathcal{B}}(H, E)$ precisely as follows:

$$\delta_{\mathcal{B}}(H,E) =_{df} \begin{cases} = +\infty & \text{if } E \vDash_{\mathcal{B}} H, \\ 0 & \text{if } E \nvDash_{\mathcal{B}} H \text{ and } E \nvDash_{\mathcal{B}} \neg H, \\ = -\infty & \text{if } E \vDash_{\mathcal{B}} \neg H \text{ and } E \nvDash_{\mathcal{B}} H. \end{cases}$$

- $\delta_{\mathcal{B}}$ is not very interesting *quantitatively*, since it can only take on *3* values. But, it will be useful for my lectures to begin with $\delta_{\mathcal{B}}$ as the simplest, most naive (classical) measure of (propositional) "degree of logical support".
- Moreover, working with $\delta_{\mathcal{B}}$ will greatly simplify certain key concepts and distinctions I will be presenting and generalizing below.
- I will begin with two such "applications" of $\delta_{\mathcal{B}}$.

Deductive Support IV: "On the Supposition That" *vs* "And"

- Consider the following two quantities:
  - The degree to which E supports H, on the supposition that (given) K.
  - The degree to which E & K supports H.
- In the classical deductive framework, we can express these two quantities formally using $\delta_{\mathcal{B}}$ (with "$\mid$" for "given that"), respectively, as follows:
  - $\delta_{\mathcal{B}}(H, E \mid K)$
  - $\delta_{\mathcal{B}}(H, E \,\&\, K)$ [this can also be expressed as $\delta_{\mathcal{B}}(H, E \,\&\, K \mid \top)$]
- But, we can say much more than this. For, in the classical deductive framework, these two quantities are in fact *identical*! That is:

(3) $$\delta_{\mathcal{B}}(H, E \mid K) = \delta_{\mathcal{B}}(H, E \,\&\, K)$$

- In classical deductive logic, saying that E entails H *on the supposition that* K just *means* that E and K are *jointly sufficient for* H (*i.e.*, that $E \,\&\, K \vDash_{\mathcal{B}} H$).

Deductive Support V: Independent *vs* Dependent Support I

- In many logic texts, a distinction is made between "independent" and "dependent" premises in arguments (first in [2], later in [9], [33, §1.6]).
- The following informal definition is typically given, and I will adopt it here.

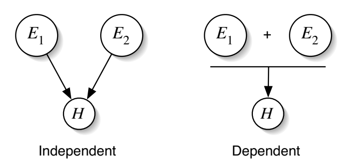
  Definition. E_1 and E_2 constitute *independent evidence* regarding H iff the degree to which E_1 supports H *on the supposition that* E_2 equals the degree to which E_1 supports H *unconditionally, and vice versa*.[a]
- In our classical deductive framework, this leads to the following notion:

  Definition. E_1 and E_2 are *deductively independent* regarding H iff $\delta_{\mathcal{B}}(H, E_1 \mid E_2) = \delta_{\mathcal{B}}(H, E_1)$ and $\delta_{\mathcal{B}}(H, E_2 \mid E_1) = \delta_{\mathcal{B}}(H, E_2)$.
- In deductive logic, this is a rather trivial notion (partly because there is no distinction between "on the supposition that" and "and"). But, there are a few things I would like to point out now, which will become very important later.

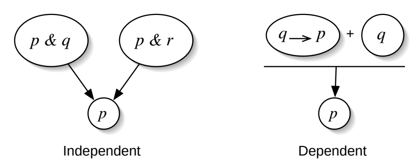
[a] There may suppositions *other than* E_1 and E_2 in a given context. We will hold these fixed.

Deductive Support VI: Independent *vs* Dependent Support II

- In logic texts, one often sees "network diagrams" like these (*e.g.*, [33, §1.6]):



- In the deductive case, the following two examples suffice to cover the cases:



- Note: $\delta_{\mathcal{B}}(p, p \,\&\, q \mid p \,\&\, r) = \delta_{\mathcal{B}}(p, p \,\&\, q \,\&\, r) = \delta_{\mathcal{B}}(p, p \,\&\, q) = \infty$. But, $\delta_{\mathcal{B}}(p, q \to p \mid q) = \delta_{\mathcal{B}}(p, (q \to p) \,\&\, q) = \infty \neq \delta_{\mathcal{B}}(p, q \to p) = \delta_{\mathcal{B}}(p, q) = 0$.

Deductive Support VII: Independent *vs* Dependent Support III

- In 1878, C.S. Peirce [47] articulated several fundamental intuitions about independent support. There, he said (my bastardization in brackets):

  > Two arguments which are ... independent, neither weakening nor strengthening the other, ought, when they concur, to produce a [degree of support] equal to the sum of the [degrees of support] which either would produce separately.

- Peirce was talking about "intensity of belief" (and not necessarily degree of support). But, let's think about what he's saying in terms of a generic measure $\mathfrak{c}$ of *degree of support* or *degree of confirmation* (an "inspirational" reading).
- On this reading, Peirce gives an *independence-additivity desideratum* for adequate measures $\mathfrak{c}$. In our framework, we can express this desideratum as:
  - ($\mathcal{A}$) If E_1 and E_2 provide independent support regarding H, according to a support measure $\mathfrak{c}$, then $\mathfrak{c}(H, E_1 \,\&\, E_2) = \mathfrak{c}(H, E_1) + \mathfrak{c}(H, E_2)$.
- How does our measure of deductive support $\delta_{\mathcal{B}}$ perform regarding $\mathcal{A}$?

Deductive Support VIII: Independent *vs* Dependent Support IV

- As it turns out, $\delta_{\mathcal{B}}$ (trivially) *satisfies* desideratum $\mathcal{A}$. Here are the cases.

$E_1 \vDash_{\mathcal{B}} H$	$E_1 \vDash_{\mathcal{B}} \neg H$	$E_2 \vDash_{\mathcal{B}} H$	$E_2 \vDash_{\mathcal{B}} \neg H$	$\delta_{\mathcal{B}}(H, E_1 \,\&\, E_2)$	$\delta_{\mathcal{B}}(H, E_1)$	$\delta_{\mathcal{B}}(H, E_2)$
True	**True**	**True**	**True**	$\boldsymbol{+\infty}$	$\boldsymbol{+\infty}$	$\boldsymbol{+\infty}$
True	**True**	**True**	**False**	$\boldsymbol{+\infty}$	$\boldsymbol{+\infty}$	$\boldsymbol{+\infty}$
True	True	False	True	$+\infty$	$+\infty$	$-\infty$
True	True	False	False	$+\infty$	$+\infty$	0
True	**False**	**True**	**True**	$\boldsymbol{+\infty}$	$\boldsymbol{+\infty}$	$\boldsymbol{+\infty}$
True	**False**	**True**	**False**	$\boldsymbol{+\infty}$	$\boldsymbol{+\infty}$	$\boldsymbol{+\infty}$
True	False	False	True	$+\infty$	$+\infty$	$-\infty$
True	False	False	False	$+\infty$	$+\infty$	0
False	True	True	True	$+\infty$	$-\infty$	$+\infty$
False	True	True	False	$+\infty$	$-\infty$	$+\infty$
False	**True**	**False**	**True**	$\boldsymbol{-\infty}$ or $+\infty$	$\boldsymbol{-\infty}$	$\boldsymbol{-\infty}$
False	True	False	False	$-\infty$ or $+\infty$	$-\infty$	0
False	False	True	True	$+\infty$	0	$+\infty$
False	False	True	False	$+\infty$	0	$+\infty$
False	False	False	True	$-\infty$ or $+\infty$	0	$-\infty$
False	**False**	**False**	**False**	**0** or $+\infty$ or $-\infty$	**0**	**0**

Deductive Support IX: Summarizing Things to This Point

- We began with the notion of (propositional, Boolean) deductive logical support. We saw that it had two particularly peculiar properties.
  - ∗ Contradictions deductively support arbitrary propositions.
  - ∗ Arbitrary propositions deductively support tautologies.
- We defined a "quantitative measure" $\delta_{\mathcal{B}}$ of "degree of deductive support in structure $\mathcal{B}$". And, we saw that it had the following peculiar property.

  $\delta_{\mathcal{B}}(H, E \mid K) = \delta_{\mathcal{B}}(H, E \,\&\, K)$

  ∴ No deductive distinction between "given that" and "and".
- We gave a precise definition of the (Peirceian) notion: "E_1 and E_2 provide *independent support* regarding H, according to measure of support $\mathfrak{c}$":

  $\mathfrak{c}(H, E_1 \mid E_2) = \mathfrak{c}(H, E_1)$ and $\mathfrak{c}(H, E_2 \mid E_1) = \mathfrak{c}(H, E_2)$

  $\Rightarrow_{\mathcal{A}}$ $\mathfrak{c}(H, E_1 \,\&\, E_2) = \mathfrak{c}(H, E_1) + \mathfrak{c}(H, E_2)$
- And, we saw that $\delta_{\mathcal{B}}$ satisfies Peirce's independence-additivity desideratum $\mathcal{A}$.

Inductive Support I: Some Historical/Motivating Remarks

- Keynes [37] and Carnap [7] were both interested in explicating an *inductive* (but still *logical*) concept of "confirmation" or "support". Though their explications ultimately failed, some of their desiderata were quite sensible.
- The basic ideas behind these early treatments of inductive support included:
  - ∗ Inductive support should be a *logical* (or, at least, *a priori* and *objective*) relation between propositions (or, if you prefer, sentences).
  - ∗ The correct explication of inductive support should involve *probability*.
  - ∗ In some sense, inductive support should be a (truly) *quantitative generalization* of the notion of *deductive* support, captured by $\delta_{\mathcal{B}}$.
- I sympathize with all three of these fundamental motivating ideas. And, I will show, below, how they can *all* be satisfied in a rather simple and intuitive way.
- But, first, I'd like to discuss some naive (and inadequate) attempts to satisfy these desiderata, including the "received view" of inductive strength.

Inductive Support II: Naive Approach #1

- As Brian Skyrms [57, ch. 2] nicely explains, one might try to define (propositional) inductive support *via* the following *naive* generalization of the definition of deductive support (here, we are still being informal):

  Naive Approach #1. E inductively supports $H \Leftrightarrow$ it is *improbable* that:
  - E is true, *and*
  - H is false.

- Skyrms [57, p. 20–21] gives two examples which show that the $\Leftarrow$ direction of this definition is not correct, intuitively. Here's the first one (note $H \vDash E$ here).
- Let E = "There is a 2000-year-old man in Cleveland," and H = "There is a 2000-year-old man in Cleveland who has three heads." As Skyrms explains:

  It is improbable for H to be false *and* E true. For H to be false and E true, there would have to be a non-three-headed 2000-year-old man in Cleveland, and it is quite improbable that that there is *any* 2000-year-old man in Cleveland. Thus, it is improbable that H is false *and* E true, *simply because it is improbable that E is true*.

Inductive Support III: Naive Approach #1, cont'd

- Let E = "There is a man in Cleveland who is 1999 years and 11-months-old and in good health," and H = "No man will live to be 2000 y.o." This time:

  ... it is improbable that E is true *and* H false, *simply because it is improbable that H is false* ... it is improbable that H is false and consequently that H is false *and* E true.

- According to Skyrms, the reason why these are not genuine cases of inductive support is that, in each case, $\Pr(\neg H \,\&\, E)$ is low, but $\Pr(\neg H \mid E)$ is *not* low. That is, (intuitively) $\neg H$ is not improbable *on the supposition that* E.
- Note: these are both examples in which $E \,\&\, \neg H$ is (intuitively) *improbable*, but *merely because either E or ¬H **alone** is improbable*. There seems to be no *relation* of support here (reminiscent of "anomalous" deductive cases!).
- While I agree with Skyrms about the inadequacy of this naive definition, I do not think he has chosen the best examples to illustrate its *insufficiency*.
- Also, Skyrms fails to mention that the improbability of $E \,\&\, \neg H$ *is* a *necessary* condition for inductive support (*even on **his own** definition*!).

Inductive Support IV: Naive Approach #2 — "The Received View" I

- Skyrms proposes an alternative approach to inductive support. His approach, which can safely be called the "Received View," is (still informally):

  Naive Approach #2. E inductively supports $H \Leftrightarrow$ it is *improbable* that H is false *given that* (or *on the supposition that*) E is true.

- At this point, we need to introduce some *probability theory* into the mix.
- A (propositional) *probability model* $\mathcal{M}$ consists of a Boolean Algebra $\mathcal{B}$ of propositions, together with a function Pr from $\mathcal{B}$ onto $[0, 1]$ such that:

Non-negativity. For all $X \in \mathcal{B}$, $\Pr(X) \geq 0$.
Normalization. For all $X \in \mathcal{B}$, if X is tautologous, then $\Pr(X) = 1$.
Additivity. For all mutually exclusive $X, Y \in \mathcal{B}$, $\Pr(X \vee Y) = \Pr(X) + \Pr(Y)$.

- And, following Kolmogorov [38] (*for now*), I *define*: $\Pr(X \mid Y) =_{df} \frac{\Pr(X \& Y)}{\Pr(Y)}$ (*if* $\Pr(Y) > 0$). By "the probability of *H given that E*" Skyrms *means* $\Pr(H \mid E)$.

Inductive Support V: Naive Approach #2 — "The Received View" II

- Now, we can more precisely compare the two naive approaches so far:
  - **Naive Approach #1**. E inductively supports $H \Leftrightarrow \Pr(\neg H \,\&\, E)$ is "low".
  - **Naive Approach #2**. E inductively supports $H \Leftrightarrow \Pr(\neg H \mid E)$ is "low".
- So, the reason Skyrms rejects the first approach is that there are cases in which (intuitively) $\Pr(\neg H \,\&\, E)$ is "low", but $\Pr(\neg H \mid E)$ is *not* "low".
- But, it is interesting to note that, for *any* probability model $\mathcal{M}$ (check this!):

$$\Pr_{\mathcal{M}}(\neg H \mid E) < \epsilon \Rightarrow \Pr_{\mathcal{M}}(\neg H \,\&\, E) < \epsilon.$$

- Therefore, "$\Pr(\neg H \,\&\, E)$ is 'low"' is a *necessary* (but *insufficient*) condition for inductive support — *according to Skyrms' own theory of inductive support*!
- In fact, I will later argue that "$\Pr(\neg H \,\&\, E)$ is 'low"' is a necessary condition for inductive support *on any reasonable* (Pr) *theory of inductive support*!
- Now, let's get a bit deeper into Skyrms' proposal (The Received View).

Inductive Support VI: Naive Approach #2 — "The Received View" III

- It seems clear to me that we can run something like Skyrms' critical arguments against *his own* definition of inductive support.
- E = "Fred Fox (who is male) takes birth control pills for 1 year," H = "At the end of that year, Fred Fox is not pregnant." Intuitively, we can say:

  It is improbable that H is false *given that* E is true, *simply because it is improbable that* H *is false* simpliciter. E *has nothing to do with it.*
- That is to say (intuitively), $\Pr(\neg H \mid E)$ is "low," *but this has nothing to do with* E. Intuitively, E is *irrelevant* to H, and for this reason E does *not support* H.
- To be a bit more precise (though still only intuitive), it seems to me that:

  $$\Pr(H \mid E) = \Pr(H)$$

  i.e., that E and H are (intuitively) *probabilistically independent*.
- It is this fact that undergirds my claim that (intuitively) E does *not* support H in the Fred Fox example. It seems that Skyrms' definition is *insufficient* ($\not\Leftarrow$).

Inductive Support VII: Naive Approach #2 — "The Received View" IV

- These considerations suggest the following (informal) *desideratum* (many authors discuss $\mathcal{R}$ — [48], [7, new preface] and [40] are early examples).

($\mathcal{R}$) E inductively supports $H \Rightarrow E$ is *not* (probabilistically) *irrelevant to* H.

- *Neither* of the naive probabilistic proposals for generalizing the notion of deductive support satisfy this *relevance* desideratum $\mathcal{R}$ (check this!).
- As we'll see below, *all* of the contemporary *Bayesian* accounts of confirmation or inductive support *do* (*a fortiori*) satisfy $\mathcal{R}$.
- Here's a more careful rendition of $\mathcal{R}$ (we'll make $\mathcal{R}$ even more precise later).

($\mathcal{R}$) E inductively *supports* H *in model* $\mathcal{M} \Rightarrow \Pr_{\mathcal{M}}(H \mid E) > \Pr_{\mathcal{M}}(H)$.
E inductively *counter*-supports H *in model* $\mathcal{M} \Rightarrow \Pr_{\mathcal{M}}(H \mid E) < \Pr_{\mathcal{M}}(H)$.
E is inductively *irrelevant* to H *in model* $\mathcal{M} \Rightarrow \Pr_{\mathcal{M}}(H \mid E) = \Pr_{\mathcal{M}}(H)$.

- Before continuing toward our general theory of support, let's backtrack briefly and consider how the naive accounts do on the three historical desiderata.

Inductive Support VIII: Keeping Score on the Historical *Desiderata* I

- Let's recall the three historical/motivating desiderata.

  (*i*) Inductive support should be a *logical* (or, at least, *a priori* and *objective*) relation between propositions (or, if you prefer, sentences).

  (*ii*) The correct explication of inductive support should involve *probability*.

  (*iii*) In some sense, inductive support should be a (truly) *quantitative generalization* of the notion of *deductive* support, captured by $\delta_{\mathcal{B}}$.
- I will be discussing (*i*) later on (Note: Skyrms' approach may fail here).
- As far as (*ii*) is concerned, both of the naive approaches use probability to explicate inductive support. So, we can't fault them on *this* score.
- Let's be a bit more precise about (*iii*). Let's require (*if* $\mathfrak{c}_{\mathcal{M}}$ is well-defined):

($\mathcal{K}$) $$\mathfrak{c}_{\mathcal{M}}(H,E) \text{ should be } \begin{cases} \textit{maximal} & \text{if } E \vDash_{\mathcal{B}} H \\ \textit{minimal} & \text{if } E \vDash_{\mathcal{B}} \neg H. \end{cases}$$

Inductive Support IX: Keeping Score on the Historical *Desiderata* II

- We can define measures of support using each of the naive ideas, as follows.
  * $\mathfrak{c}^1_{\mathcal{M}}(H,E) =_{df} 1 - \Pr_{\mathcal{M}}(\neg H \,\&\, E)$.
  * $\mathfrak{c}^2_{\mathcal{M}}(H,E) =_{df} 1 - \Pr_{\mathcal{M}}(\neg H \mid E) = \Pr_{\mathcal{M}}(H \mid E)$.
- Now, let's see how $\mathfrak{c}^1_{\mathcal{M}}$ and $\mathfrak{c}^2_{\mathcal{M}}$ do with respect to $\mathcal{K}$ (*if* they're well-defined).

$$\mathfrak{c}^1_{\mathcal{M}}(H,E) = \begin{cases} 1 \text{ (maximal)} & \text{if } E \vDash_{\mathcal{B}} H \\ \Pr_{\mathcal{M}}(\neg E) \text{ (\textit{not} minimal)} & \text{if } E \vDash_{\mathcal{B}} \neg H. \end{cases}$$

$$\mathfrak{c}^2_{\mathcal{M}}(H,E) = \begin{cases} 1 \text{ (maximal)} & \text{if } E \vDash_{\mathcal{B}} H \\ 0 \text{ (minimal)} & \text{if } E \vDash_{\mathcal{B}} \neg H. \end{cases}$$

- This exposes another deficiency in the first naive approach to inductive support. The first naive approach does *not* generate a measure which properly *generalizes* $\delta_{\mathcal{B}}$. On the other hand, The Received View *does* satisfy (*iii*).
- Next, we inch closer to a general, rigorous probabilistic theory of support . . .

Inductive Support X: Back to "Given That," "And," and "Independence" I

- In probability theory, there is a *big* difference between "given that" and "and". As a result, probabilistic measures of support $\mathfrak{c}$ will generally be such that

$$\mathfrak{c}(H, E \mid K) \neq \mathfrak{c}(H, E \,\&\, K).$$

- $\mathfrak{c}_{\mathcal{M}}(H, E \mid K)$ denotes **the degree to which E inductively supports H given that (on the supposition that) K, in probability model $\mathcal{M}$, according to measure $\mathfrak{c}$**. This is (inherently — see below) a *five-place* function.
- In general, $\mathfrak{c}$ will, in turn, be defined in terms of the probability function $\Pr_{\mathcal{M}}$ *specified in the model* $\mathcal{M}$. We obtain $\mathfrak{c}_{\mathcal{M}}(H, E \mid K)$ from $\mathfrak{c}_{\mathcal{M}}(H, E)$ *via conditionalizing* all occurrences of $\Pr_{\mathcal{M}}$ in the definition of $\mathfrak{c}_{\mathcal{M}}$ on K.
- **Crucial Note**: You *cannot* think of a subscripted $\mathcal{M}$ in $\mathfrak{c}_{\mathcal{M}}$ or $\Pr_{\mathcal{M}}$ as involving *conditionalization* on $\mathcal{M}$ ($\mathcal{M}$ is merely an *index*!). This is because:

$$\Pr_{\mathcal{M}}(X) \neq \Pr_{?}(X \mid \mathcal{M}).$$

- $\mathcal{M} \notin \mathcal{B}$! "$\Pr(X \mid \mathcal{M})$" is *nonsense*! $\mathcal{M}$ *indexes* the probability function $\Pr_{\mathcal{M}}$.

Inductive Support XI: Back to "Given That," "And," and "Independence" II

- Now, we're ready to make our relevance desideratum $\mathcal{R}$ *fully* general and *quantitative*. $\mathcal{R}$ is best understood as a desideratum on *measures* of support:

$$(\mathcal{R}) \qquad \mathfrak{c}_{\mathcal{M}}(H, E \mid K) \text{ should be } \begin{cases} > 0 & \text{if } \Pr_{\mathcal{M}}(H \mid E \,\&\, K) > \Pr_{\mathcal{M}}(H \mid K) \\ 0 & \text{if } \Pr_{\mathcal{M}}(H \mid E \,\&\, K) = \Pr_{\mathcal{M}}(H \mid K) \\ < 0 & \text{if } \Pr_{\mathcal{M}}(H \mid E \,\&\, K) < \Pr_{\mathcal{M}}(H \mid K) \end{cases}$$

- Measures satisfying $\mathcal{R}$ are called *relevance measures*. We will see *many* relevance measures in our discussion of Bayesian confirmation, below.
- We can also now give a fully rigorous definition of "E_1 and E_2 provide *independent support regarding H* in a model $\mathcal{M}$, according to measure $\mathfrak{c}_{\mathcal{M}}$."

  Definition. E_1 and E_2 provide *independent support regarding H* in a probability model $\mathcal{M}$, according to a measure of support $\mathfrak{c}_{\mathcal{M}}$ iff $\mathfrak{c}_{\mathcal{M}}(H, E_1 \mid E_2) = \mathfrak{c}_{\mathcal{M}}(H, E_1)$ and $\mathfrak{c}_{\mathcal{M}}(H, E_2 \mid E_1) = \mathfrak{c}_{\mathcal{M}}(H, E_2)$.[a]

[a] There may suppositions *other than* E_1 and E_2 in a given context. These are assumed to be *held fixed* in independence judgments concerning E_1 and E_2. We'll see examples of such judgments, below.

Inductive Support XII: A *General* Measure of Support

- Consider the following relevance measure of the degree to which E supports H given that (on the supposition that) K, in probability model $\mathcal{M}$:

$$l_{\mathcal{M}}(H, E \mid K) =_{df} \log\left[\frac{\Pr_{\mathcal{M}}(E \mid H \& K)}{\Pr_{\mathcal{M}}(E \mid \neg H \& K)}\right].$$

- The measure $l_{\mathcal{M}}$ is called the *log-likelihood-ratio* measure. $l_{\mathcal{M}}$ has been used and defended by several authors (see [26], [24], [21] and references therein).
- We'll see below that $l_{\mathcal{M}}$ has *many* desirable (and unique) properties (relative to other relevance measures). For now, note that (exercises: prove these!):
  * $l_{\mathcal{M}}$ satisfies *all* the historical desiderata (*including* the *deductive* $\mathcal{K}$).
  * $l_{\mathcal{M}}$ satisfies the relevance desideratum $\mathcal{R}$.
  * $l_{\mathcal{M}}$ satisfies Peirce's independence-additivity desideratum $\mathcal{A}$.
  * $l_{\mathcal{M}}$ is *silent* (undefined) on *both* "anomalous" deductive cases.
- I'll return to some of these later. Next, I'll discuss Bayesian confirmation. But, first, a sidebar on subjectivity, objectivity, logic, and probability.

Sidebar: Subjectivity, Objectivity, Logical Support, and Probability I

- Carnap (early on) thought that providing a *logical* theory of support required explaining how *logic alone* can determine probability values $\Pr_{\mathcal{L}}(X \mid Y)$.
- Carnap (initially) *identified* the degree to which E confirms (*logically supports*) H given that K with $\Pr_{\mathcal{L}}(H \mid E \,\&\, K)$, where $\Pr_{\mathcal{L}}$ is "logical."
- He spent most of his life trying to explicate the notion of "logical probability" $\Pr_{\mathcal{L}}(X \mid Y)$ ("probabilities determined by *logical structure alone*").
- The idea was that if we could determine the values of $\Pr_{\mathcal{L}}(H \mid \top)$, then we could determine the degree of support provided for H *any conjunction of propositions*, simply by *conditionalizing* $\Pr_{\mathcal{L}}(H \mid \top)$ on them.
- He never succeeded in explicating his $\Pr_{\mathcal{L}}(H \mid \top)$. The consensus now seems to be that this is all but hopeless (see [41], [42], [15] and [3] for discussion).
- I think this is a *category error*. In the Boolean case, there's no such thing as "$\Pr_{\mathcal{B}}(H)$". The *deductive* structure $\mathcal{B}$ contains *no* probabilistic information!

Sidebar: Subjectivity, Objectivity, Logical Support, and Probability II

- In order to determine *probabilities*, one needs an *inductive* structure. That is, one needs a *probability model* $\mathcal{M}$. A *deductive* structure $\mathcal{L}$ isn't enough.
- A theory of inductive logic need not provide a recipe for producing probabilities from deductive structure alone. Indeed, there is no such recipe!
- The theory of inductive support I'm sketching is still *objective* (and if you like *a priori*). Once a probability model $\mathcal{M}$ is specified, the relations of inductive support — *in* $\mathcal{M}$ — are *objectively* (*deductively*!) determined.
- At this point, you probably want to ask: "But, where do the *models* $\mathcal{M}$ come from?" This is a good question, but is it a *logical* question?
- It needn't be the business of inductive *logic* to tell us which probability models we should use. The *logic* of inductive support is *model-relational*.
- Some models may involve "subjective" elements (see below), but this doesn't make the *relations within those models* subjective or *a posteriori*. To wit ...

Bayesian Confirmation I: Some Background I

- Bayesianism (*i.e.*, Bayesian *epistemology*) assumes that the degrees of belief (or credence) of rational agents are *probabilities* (for now, Kolmogorov [38]).
- $\Pr_{\mathcal{M}_a}(H)$ denotes the unconditional probability of H *in the probability model set-up by the rational Bayesian agent a's degree of belief structure*.
- $\Pr_{\mathcal{M}_a}(H \mid E)$ denotes the conditional probability of H given that (or *on a's supposition that*) E *in the probability model set-up by the rational Bayesian agent a's degree of belief structure*. Note: this can all be *synchronic*.
- There is much controversy over whether rational degrees of belief really are probabilities, and over the objective status and origins of *prior* probabilities (or the *models* $\mathcal{M}_a$!). For now, all such questions will be bracketed.
- To simplify and focus the discussion (for now), I will assume (*arguendo!*) that *all* rational Bayesian agents share a *single* probability function (and model) Pr.

Bayesian Confirmation II: Some Background II

- In Bayesian confirmation theory, evidence E *confirms* (or *supports*) a hypothesis H (given K) if E and H are positively correlated under $\Pr(\cdot \mid K)$.
- If E and H are negatively correlated under $\Pr(\cdot \mid K)$, then E *dis*confirms (or *counter*-supports) H (given K), and if E and H are *independent* under $\Pr(\cdot \mid K)$, then E is confirmationally *neutral* regarding H (given K).
- Within (Kolmogorov! [16]) probability theory, there are many *logically equivalent* ways of saying that E confirms H (given K). Here are a few:
  - ∗ E confirms H (given K) if $\Pr(H \mid E \,\&\, K) > \Pr(H \mid K)$.
  - ∗ E confirms H (given K) if $\Pr(E \mid H \,\&\, K) > \Pr(E \mid \neg H \,\&\, K)$.
  - ∗ E confirms H (given K) if $\Pr(H \mid E \,\&\, K) > \Pr(H \mid \neg E \,\&\, K)$.
- By taking differences, ratios, *etc.*, of the left/right sides of these alternative inequalities, a *plethora* of possible *quantitative (relevance) measures* of the *degree* to which E confirms (or supports) H can be formed.

Bayesian Confirmation III: Four Popular/Representative Contemporary Measures

- *Dozens* of Bayesian relevance measures have been proposed in the philosophical literature (see [39] for a survey). Here are four popular ones.[a]
  - ∗ *Difference*: $d(H, E \mid K) =_{df} \Pr(H \mid E \,\&\, K) - \Pr(H \mid K)$
  - ∗ Log-*Ratio*: $r(H, E \mid K) =_{df} \log\left[\frac{\Pr(H \mid E \,\&\, K)}{\Pr(H \mid K)}\right]$
  - ∗ Log-*Likelihood-Ratio*: $l(H, E \mid K) =_{df} \log\left[\frac{\Pr(E \mid H \,\&\, K)}{\Pr(E \mid \neg H \,\&\, K)}\right]$
  - ∗ "Normalized" d: $s(H, E \mid K) =_{df} \Pr(H \mid E \,\&\, K) - \Pr(H \mid \neg E \,\&\, K) = \frac{1}{\Pr(\neg E \mid K)} \cdot d(H, E \mid K)$
- Logs are taken to ensure easy satisfaction of desiderata $\mathcal{R}$ and $\mathcal{A}$ (how?). They are merely a useful convention (they're inessential, but they simplify things).
- The first part of our story concerns the *disagreement* exhibited by these measures, and its ramifications for Bayesian confirmation theory ...

[a] Users of d include [12], [11], and [34]. Users of r include [31], [43], and [32]. Users of l include [26], [56], and [21]. Users of s include [35] and [8]. See [16], [19], and [21] for further references.

Bayesian Confirmation IV: Disagreement Between Alternative Measures

- What kind of disagreement between relevance measures is important?
- Mere *numerical* (or *conventional* or *syntactical*) differences between measures are not important, since they need not effect *ordinal* judgments of what is more/less well confirmed than what (by what).
- *Ordinal* differences are crucial, since they can effect the cogency of many arguments surrounding Bayesian confirmation theory.
- For instance, it is part of Bayesian lore that the observation of a black raven (E_1) confirms the hypothesis (H) that all ravens are black *more strongly than* the observation of a white shoe (E_2) does (given "actual corpus" K).
- But, given the standard background assumptions (K) in Bayesian accounts of Hempel's ravens paradox, this conclusion [$\mathfrak{c}(H, E_1 \mid K) > \mathfrak{c}(H, E_2 \mid K)$] follows only for *some* measures of confirmation $\mathfrak{c}$ (and *not* others).
- Such arguments are said to be *sensitive to choice of measure* [19].

Bayesian Confirmation V: The Problem of Measure Sensitivity

- A detailed study of the literature shows that *virtually every argument* involving quantitative Bayesian confirmation theory is sensitive to choice of measure [19]! Below, I'll discuss the measure sensitivity of the following:
  * The Popper-Miller Argument *Against* Bayesianism
  * Rosenkrantz and Earman on the Problem of "Irrelevant Conjunction"
  * Eells and Sober on Goodman's "Grue" Paradox
  * Horwich *et al.* on Hempel's Ravens Paradox
  * Horwich *et al.* on the Confirmational Value of Varied Evidence
  * Earman on the problem of old evidence
- There are many other important measure-sensitive arguments [6], [4], [5].
- One needn't gerrymander or comb the historical literature for Bayesian relevance measures which fail to undergird these arguments.
- Each of these arguments is valid with respect to *only some* of d, r, l, and s.

Bayesian Confirmation VI: The Popper-Miller Argument *Against* Bayesianism

- It isn't just arguments/accounts *within* Bayesian confirmation theory that are sensitive to choice of measure. Some well-known *criticisms* of Bayesianism also rest on measure sensitive arguments.
- Most famously, Popper and Miller ([50], [22]) use the following property of the difference measure d to argue *against* Bayesianism (generally):

(4) $$d(H, E) = d(H \vee E, E) + d(H \vee \neg E, E).$$

- As it turns out, neither the log-ratio measure r [51], nor the log-likelihood-ratio measure l [27] satisfies property (4) (check this!).
- $\therefore$ The Popper-Miller argument is *sensitive to choice of measure*.
- In the absence of reasons to think that d is a more accurate (and charitable) reconstruction of Bayesian confirmation theory than either r or l, the Popper-Miller argument remains (at best) *enthymematic*.

Bayesian Confirmation VII: Rosenkrantz on "Irrelevant Conjunction"

- Rosenkrantz [55] provides a Bayesian resolution of the problem of Irrelevant Conjunction (*a.k.a.*, the Tacking Problem) which trades on the following property of the difference measure:

(5) $$\text{If } H \vDash E, \text{ then } d(H \,\&\, X, E) = \Pr(X \mid H) \cdot d(H, E).$$

- Neither r nor l satisfies property (5) [17].
- Rosenkrantz does provide some (pretty good) reasons to reject r. However, he [54] explicitly admits that he knows of "no compelling considerations that adjudicate between" d and l.
- So, it is (at best) unclear how one might consistently complete Rosenkrantz's enthymematic treatment of the tacking problem.
- What's worse, as I will explain later, I think there are good reasons to *favor l over d* as a measure of support.

Bayesian Confirmation VIII: Earman on "Irrelevant Conjunction"

- Earman [11] gives a more robust resolution of the tacking problem which requires only the following logically weaker cousin of (5):

(5′) If $H \vDash E$, then $d(H \mathbin{\&} X, E) < d(H, E)$.

- r violates even this weaker condition, but l *satisfies* (5′) [17].
- In this sense, Earman's account is *less* sensitive to choice of measure (*i.e.*, more robust) than Rosenkrantz's is.
- Earman's account can be bolstered by providing compelling independent reasons to favor d (or l) over r (*e.g.*, see below).
- Unfortunately, even the bolstered version of Earman's account is inadequate. I provide a new and improved Bayesian resolution of the problem of irrelevant conjunction in [17].

Bayesian Confirmation IX: Eells on Goodman's "Grue" Paradox

- Eells [12] offers a Bayesian account of the Grue paradox (*a.k.a.*, Goodman's "new riddle of induction") which trades on the following property of the difference measure [where β, δ are: $\beta =_{df} \Pr(H_1 \mathbin{\&} E) - \Pr(H_2 \mathbin{\&} E)$, and $\delta =_{df} \Pr(H_1 \mathbin{\&} \neg E) - \Pr(H_2 \mathbin{\&} \neg E)$]:

(6) If $\beta > \delta$ and $\Pr(E) < \frac{1}{2}$, then $d(H_1, E) > d(H_2, E)$.

- Neither r nor l satisfies property (6).
- Eells does provide reasons (as reported in a paper by Sober, see below) to prefer the difference measure d over the log-ratio measure r, but he does not supply reasons to prefer d over l.
- Pending such reasons, Eells's argument remains *enthymematic*.
- Moreover, I will later provide reasons to favor l over d.

Bayesian Confirmation X: Sober on Goodman's "Grue" Paradox

- Sober [59] describes a more robust Bayesian account of the Grue paradox which exploits the following weaker property of d:

(6′) If H_1, H_2 entail E and $\Pr(H_1) > \Pr(H_2)$, then $d(H_1, E) > d(H_2, E)$.

- r violates even this weaker condition, but l *satisfies* (6′).
- In this sense, Sober's resolution of Goodman's "Grue" paradox is *less* sensitive to choice of measure (*i.e.*, more robust) than Eells's is.
- And, like Eells, Sober does provide *some* reasons to prefer d to r.
- However, as I explain in my [19] and [21], these reasons (which are borrowed from Eells) are not very good reasons to prefer d to r.
- Like Earman's account of "Irrelevant Conjunction," Sober's account of "Grue" can be bolstered by providing compelling independent reasons to favor d (or l) over r (*e.g.*, see below).

Bayesian Confirmation XI: Horwich *et al.* on Ravens & Variety of Evidence

- The vast majority of Bayesian explications of both the Ravens Paradox and the confirmational value of varied evidence presuppose the following ($\mathfrak{c}$ ranges over individual measures):

(7) If $\Pr(H \mid E_1) > \Pr(H \mid E_2)$, then $\mathfrak{c}(H, E_1) > \mathfrak{c}(H, E_2)$.

- The "normalized" difference measure s *violates* (7).[a]
- Typically, the advocates of such arguments have used either d or r in their arguments (as it turns out, d, r, and l *all* satisfy (7)).
- None of these authors seems to provide (*independent*) reasons to prefer their measures over s (or other measures which violate (7)).
- In my [20] and [21], I propose a novel Bayesian explication of the confirmational value of varied evidence, based on l (see below).

[a] So do Carnap's [7, §67] relevance measure $\mathfrak{r}(H, E) = \Pr(H \mathbin{\&} E) - \Pr(H) \cdot \Pr(E)$, Mortimer's [44] measure $\Pr(E \mid H) - \Pr(E)$, and Nozick's [45] measure $\Pr(E \mid H) - \Pr(E \mid \neg H)$.

Bayesian Confirmation XII: Earman on the Problem of Old Evidence

- Earman [11, pp. 120–121] argues that quantitative Bayesian confirmation theory, together with the "radical probabilism" of Jeffrey [34] does not suffice to avoid Glymour's problem of old evidence [23, pp. 63–69].
- His argument presupposes that Bayesians use d to measure degree of confirmation, and it rests on the following fact about d:

(8) If $H \vDash E$, then $\Pr(E) \approx 1 \Rightarrow d(H, E) \approx 0$.

- This argument has two flaws. First, (8) does hold for d and r, but it does *not* hold for l or s (contrary to what Earman suggests [11, p. 243, note 8]). Second, this argument only applies to the case of *deductive evidence* ($H \vDash E$).
- As it turns out, we can avoid Earman's objections, by using our l instead of d:

(9) *Even if* $H \vDash E$ *and* $\Pr(E) \approx 1$, $l(H, E)$ can be *arbitrarily large*.

- As Joyce [35] and Christensen [8] point out, s *also* satisfies (9).

Bayesian Confirmation XIII: Attempts to Resolve the Measure-Sensitivity Problem

- There do exist a few general arguments in the literature which aim to rule-out all but a small class of ordinally equivalent measures (*e.g.*, Milne [43], Good [25], Carnap [7], Kemeny & Oppenheim [36], and Heckerman [29]).
- Others have given "piecemeal" arguments which attack a *particular* class of measures, but fail to rule-out other competing measures (*e.g.*, Rosenkrantz [55], Earman [11], Gillies [22], Eells, and Sober [59]).
- In my dissertation [21], I provide a thorough survey of both kinds of arguments, and I show that none of them is completely satisfactory.
- Most notably, I have seen (in the literature[a]) *no* compelling reasons to prefer the difference measure d over either l or s.
- Until such reasons are provided, the arguments of Gillies, Rosenkrantz, Eells, Horwich *et al.* will remain *enthymematic*.

[a] Recent joint work of Eells & Fitelson in [13] and [14] has filled this gap in the literature.

Bayesian Confirmation XIV: Tabular Summary of Key Results — So Far

	Valid *wrt* relevance measure:			
Argument/Desideratum	d?	r?	l?	s?
Deductive Generalization Desideratum $\mathcal{K}$	No	No	Yes	No
Peirceian Independence-Additivity Desideratum $\mathcal{A}$	Yes	Yes	Yes	No
Relevance Desideratum $\mathcal{R}$	Yes	Yes	Yes	Yes
Rosenkrantz on Irrelevant Conjunction	Yes	No	No	Yes
Earman on Irrelevant Conjunction	Yes	No	Yes	Yes
Eells on the Grue Paradox	Yes	No	No	Yes
Sober on the Grue Paradox	Yes	No	Yes	Yes
Horwich *et al.* on Ravens & Variety	Yes	Yes	Yes	No
Popper-Miller's *Critique* of Bayesianism	Yes	No	No	Yes
Earman's Old Evidence *Critique* of Bayesianism	Yes	Yes	No	No

Inductive Desiderata I: Symmetries and Asymmetries in Evidential Support[a]

- Consider the following two propositions concerning a card c, drawn at random from a standard deck of playing cards (classical model $\mathcal{M}$):

E: c is the ace of spades. H: c is *some* spade.

- I take it as intuitively clear and uncontroversial that:
  1. The degree to which E supports $H \neq$ the degree to which H supports E, since $E \vDash H$, but $H \nvDash E$. Intuitively, we have $\mathfrak{c}(H, E) \gg \mathfrak{c}(E, H)$.
  2. The degree to which E confirms $H \neq$ the degree to which $\neg E$ *dis*confirms H, since $E \vDash H$, but $\neg E \nvDash \neg H$. Intuitively, $\mathfrak{c}(H, E) \gg -\mathfrak{c}(H, \neg E)$.
- Therefore, *no adequate relevance measure of support* $\mathfrak{c}$ *should be such that either* $\mathfrak{c}(H, E) = \mathfrak{c}(E, H)$ *or* $\mathfrak{c}(H, E) = -\mathfrak{c}(H, \neg E)$ (for all E and H and all models $\mathcal{M}$). I'll call these two symmetry desiderata $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively.
- Note: for all H, E, and $\mathcal{M}$, $r(H, E) = r(E, H)$ and $s(H, E) = -s(H, \neg E)$. That is, r violates $\mathcal{S}_1$ and s violates $\mathcal{S}_2$. *Both* d and l satisfy *both* $\mathcal{S}$-desiderata.

[a] This slide is drawn from recent joint work of Eells & Fitelson [14].

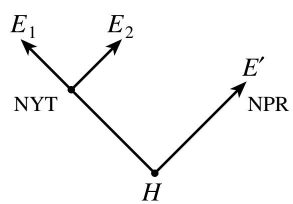
Inductive Desiderata II: Independent Evidence (Again) I

- Wittgenstein [60] alludes to a man who is doubtful about the reliability of a story he reads in the newspaper, so he buys another copy of the same issue of the same newspaper to corroborate (*not* a good strategy).
- To fix our ideas, let's assume that the story in the NYT reports that (H) the Yankees won the world series. Let E_n be the evidence obtained by reading the n^{th} copy of the same issue of the NYT.
- Intuitively, the degree to which E_2 confirms H *depends on whether* E_1 *has already been observed* (or is *supposed*). [$\mathfrak{c}(H, E_2 \mid E_1) \ll \mathfrak{c}(H, E_2)$]
- But, an *independently derived* report E' (say, one heard on a NPR broadcast) *would* corroborate the NYT story [$\mathfrak{c}(H, E' \mid E_1) = \mathfrak{c}(H, E')$].
- Intuitively, $\mathfrak{c}(H, E_1 \,\&\, E') = \mathfrak{c}(H, E_1) + \mathfrak{c}(H, E') > \mathfrak{c}(H, E_1)$, whereas $\mathfrak{c}(H, E_1 \,\&\, E_2) = \mathfrak{c}(H, E_1) = \mathfrak{c}(H, E_2)$. This is just Peirce's $\mathcal{A}$ again.
- Wittgenstein's story is an *inductive* example of *dependent support*.

Inductive Desiderata II: Independent Evidence (Again) II

- How can we *probabilistically* explain the epistemic difference between these two examples? Intuitively, a NYT report (E) and a NPR report (E') are *independent* in a way that two NYT reports (E_1, E_2) are not.
- It is *not* that the NYT report and the NPR report are (probabilistically) independent *unconditionally*, since hearing the NPR report that H or $\neg H$ raises the probability that the NYT will also report that H (or $\neg H$).
- As Sober [58] explains, the relevant fact is that E and E' are independent — *given that* H (*or* $\neg H$). That is, *once we learn* (*or suppose*) *the truth-value of* H, observing E' no longer raises the probability of E (in our intuitive $\mathcal{M}$).
- When this happens, we say that H *screens-off* E from E'.
- Learning that someone has a fever raises the probability that they have a cough, but *not* if it's already known that they have the flu. As Reichenbach [52] taught us: *common causes screen-off their joint effects from each other*.

Inductive Desiderata II: Independent Evidence (Again) III



- H does *not* screen-off E_1 from E_2 (although perhaps the state of the NYT printing press just prior to printing *does* screen off E_1 from E_2).
- On the other hand (provided that there was no communication between NYT and NPR, *etc.*), H *does* screen-off E_i from E'.
- This is the sense in which E_1 and E_2 do *not* provide *independent* support for H (although, perhaps they provide independent evidence about the state of the NYT printing press!), while E_i and E' *do*.

Inductive Desiderata II: Independent Evidence (Again) IV

- The figure in the previous slide should remind you of the "deductive networks" we saw in our "deductive" discussion of *independent evidence*.
- Such causal/probabilistic diagrams have become *famously* known as "Bayesian Networks" [46], [29], [30], [5].
- We write "$E_1 \perp\!\!\!\perp E_2 \mid H$" as shorthand for "$E_1$ and E_2 are independent, *given that* H" (H *screens-off* E_1 from E_2). More formally, we have:

  $E_1 \perp\!\!\!\perp E_2 \mid H$ iff $\Pr(E_1 \mid E_2 \,\&\, \pm H) = \Pr(E_1 \mid \pm H)$.
- Our discussion of Wittgenstein's example suggests the following desideratum for relevance measures of support $\mathfrak{c}$, and provides the key to unraveling the concept of independent inductive support.

  ($\mathcal{I}$) $E_1 \perp\!\!\!\perp E_2 \mid H \Rightarrow \mathfrak{c}(H, E_1 \mid E_2) = \mathfrak{c}(H, E_1)$ and $\mathfrak{c}(H, E_2 \mid E_1) = \mathfrak{c}(H, E_2)$.
- The only relevance measure that satisfies $\mathcal{I}$ is l [20], [29].

Postscript I: Kolmogorov, Popper, and Confirmation I

- I have been assuming Kolmogorov's [38] definition of $\Pr(X \mid Y)$. Historically, several philosophers have *not* done so (and for serious reasons).
- Most notably, Carnap [7] and Popper [49] take conditional probability as primitive and define unconditional probability in terms of *it*.
- More recently, our own Alan Hájek [28] has argued quite forcefully that the Kolmogorov definition of $\Pr(X \mid Y)$ is inadequate in many respects.
- Popper's definition of $\Pr(X \mid Y)$ is in many important ways the most general (see [53] for *encyclopedic* discussion about conditional probability).
- Popper functions and Kolmogorov functions disagree about $\Pr(X \mid Y)$ *only* in cases where $\Pr(Y \mid \top) = 0$. In these cases, Kolmogorov functions are undefined, but Popper functions are perfectly well-defined.
- This difference may seem insignificant, but it is actually *very* important, and has *serious* ramifications for Bayesian Confirmation Theory.

Postscript I: Kolmogorov, Popper, and Confirmation II

- Recall that, on the Kolmogorov theory of Pr, there are many *logically equivalent* ways to say that E and H are correlated under Pr.
- This lead to many *logically equivalent* ways of defining the *qualitative* relation of Bayesian confirmation (although, as we have seen, this also lead to many *non*-equivalent *quantitative* Bayesian *measures* of confirmation).
- On the Popper theory, the following inequalities are *not* equivalent (why?).
  - $\Pr(H \mid E) > \Pr(H \mid \top)$.
  - $\Pr(E \mid H) > \Pr(E \mid \neg H)$.
  - $\Pr(H \mid E) > \Pr(H \mid \neg E)$.
- This means that, if one adopts Popper's theory of Pr, one ends-up with many distinct notions of "correlation" or "independence", and one also ends-up with *many* notions of Bayesian confirmation! Quite a disunifying effect!

Postscript I: Kolmogorov, Popper, and Confirmation III

- This has not been widely discussed (see [16] for a notable exception).
- Unfortunately, several authors who have recently applied Popper functions in Bayesian confirmation theory do not seem to be aware of this issue.
- Most notably, Joyce [35] and Christensen [8] both think that using Popper functions is the key to solving the problem of old evidence.
- In cases of old evidence, $\Pr(E \mid \top) = 1$. Hence, on the Kolmogorov-Bayesian theory of confirmation, E *cannot* confirm H (prove this!).
- However, on the Popper-Bayesian theory, E *can* still confirm H — in *one* of the *many* Popper-Bayesian senses — because $\Pr(H \mid E)$ *can* be greater than $\Pr(H \mid \neg E)$ in the Popper theory of Pr *even if* $\Pr(E \mid \top) = 1$.
- Note: if one defines "confirmation" as $\Pr(H \mid E) > \Pr(H \mid \top)$, then one *still* has a problem of old evidence — *even in Popper-Bayes* [18] (prove this!).

Postscript II: Naive Approach #1 — Revisited

- As I mentioned in a previous slide, $\Pr(\neg H \,\&\, E) < \epsilon$ is a *necessary condition* for $\Pr(\neg H \mid E) < \epsilon$. As it turns out, $\Pr(\neg H \,\&\, E) < \epsilon$ is a much more general necessary condition than this. The following are all theorems (check these!).
  1. $d(H, E) > 1 - \epsilon \Rightarrow \Pr(\neg H \,\&\, E) < \epsilon$
  2. $r(H, E) > n > 1 \Rightarrow \Pr(\neg H \,\&\, E) < \frac{1}{n}$
  3. $l(H, E) > n > 1 \Rightarrow \Pr(\neg H \,\&\, E) < \frac{1}{n}$
  4. $s(H, E) > 1 - \epsilon \Rightarrow \Pr(\neg H \,\&\, E) < \epsilon$
- In other words, $\Pr(\neg H \,\&\, E) < \epsilon$ is a necessary condition for "E provides *strong* evidence in favor of H" — on just about any (probabilistic) theory of strong evidence that anyone has ever proposed (or is likely to propose).
- In this sense, the first Naive Approach is superior to the second ($\Pr(H \mid E) < \epsilon$ is *not* a necessary condition for *strong* evidence in the *relevance* sense!). Naive Approach #1 is more interesting than [57, ch. 2] would have us believe!

Postscript III: The Four Relevance Measures *as Generalizations of* $\delta_{\mathcal{B}}$ (Details)

- $l(H,E) = \begin{cases} +\infty & \text{if } E \vDash H, \Pr(E) > 0, \Pr(H) \in (0,1) \\ 0 & \text{if } E \perp\!\!\!\perp H, \Pr(E) > 0, \Pr(H) \in (0,1) \\ -\infty & \text{if } E \vDash \neg H, \Pr(E) > 0, \Pr(H) \in (0,1) \end{cases}$
- $d(H,E) = \begin{cases} \Pr(\neg H) & \text{if } E \vDash H, \Pr(E) > 0 \\ 0 & \text{if } E \perp\!\!\!\perp H, \Pr(E) > 0 \\ -\Pr(H) & \text{if } E \vDash \neg H, \Pr(E) > 0 \end{cases}$
- $r(H,E) = \begin{cases} \log\left[\frac{1}{\Pr(H)}\right] & \text{if } E \vDash H, \Pr(E) > 0, \Pr(H) > 0 \\ 0 & \text{if } E \perp\!\!\!\perp H, \Pr(E) > 0, \Pr(H) > 0 \\ -\infty & \text{if } E \vDash \neg H, \Pr(E) > 0, \Pr(H) > 0 \end{cases}$
- $s(H,E) = \begin{cases} \Pr(\neg H \mid \neg E) & \text{if } E \vDash H, \Pr(E) \in (0,1) \\ 0 & \text{if } E \perp\!\!\!\perp H, \Pr(E) \in (0,1) \\ -\Pr(H \mid \neg E) & \text{if } E \vDash \neg H, \Pr(E) \in (0,1) \end{cases}$

References

[1] A. R Anderson and N. D. Belnap, Jr., *Entailment*, Princeton University Press, 1975.

[2] M.C. Beardsley, *Practical logic*, Prentice Hall, 1950.

[3] J.M. Bernardo and A.F.M. Smith, *Bayesian theory*, 2 ed., John Wiley & Sons Ltd., 2000.

[4] L. Bovens, B. Fitelson, S. Hartmann, and J. Snyder, *Too odd (not) to be true? A reply to Erik J. Olsson*, *British Journal for the Phil. of Sci.* (forthcoming) `http://fitelson.org/bjps.pdf`.

[5] L. Bovens and S. Hartmann, *Bayesian networks and the problem of unreliable instruments*, Philosophy of Science **69** (2002), 29–72.

[6] D. Bradley and B. Fitelson, *Monty hall, doomsday, and confirmation*, *Analysis* (forthcoming) `http://fitelson.org/monty.pdf`.

[7] R. Carnap, *Logical foundations of probability*, second ed., University of Chicago Press, 1962.

[8] D. Christensen, *Measuring confirmation*, Journal of Philosophy **XCVI** (1999), 437–61.

[9] I.M. Copi and C. Cohen, *Introduction to logic*, eleventh ed., Prentice Hall, 2001.

[10] M. Dummett, *Elements of intuitionism*, second ed., Oxford University Press, 2000.

[11] J. Earman, *Bayes or bust?*, MIT Press, 1992.

[12] E. Eells, *Rational decision and causality*, Cambridge University Press, 1982.

[13] E. Eells and B. Fitelson, *Measuring confirmation and evidence*, Journal of Philosophy **XCVII** (2000), no. 12, 663–672, `http://fitelson.org/dccfin.pdf`.

[14] ______, *Symmetries and asymmetries in evidential support*, Philosophical Studies **107** (2002), no. 2, 129–142, `http://fitelson.org/symmetry.pdf`.

[15] R. Festa, *Optimum inductive methods*, Kluwer Academic Publishers, 1993.

[16] ______, *Bayesian confirmation*, Experience, Reality, and Scientific Explanation (M. Galavotti and A. Pagnini, eds.), Kluwer Academic Publishers, 1999, pp. 55–87.

[17] B. Fitelson, *Putting the irrelevance back into the problem of irrelevant conjunction*, forthcoming in *Philosophy of Science*, `http://fitelson.org/ic.pdf`.

[18] ______, *Review of Joyce's [35]*, *Mind* (forthcoming).

[19] ______, *The plurality of Bayesian measures of confirmation and the problem of measure sensitivity*, Philosophy of Science **66** (1999), S362–S378, `http://fitelson.org/psa.pdf`.

[20] ______, *A Bayesian account of independent evidence with applications*, Philosophy of Science **68** (2001), no. 3, suppl., S123–S140, `http://fitelson.org/psa2.pdf`.

[21] ______, *Studies in Bayesian confirmation theory*, Ph.D. thesis, University of Wisconsin–Madison (Philosophy), 2001, `http://fitelson.org/thesis.pdf`.

[22] D. Gillies, *In defense of the Popper-Miller argument*, Philosophy of Science **53** (1986), 110–113.

[23] C. Glymour, *Theory and evidence*, MIT Press, 1980.

[24] I. Good, *Good thinking: The foundations of probability and its applications*, University of Minnesota Press, 1983.

[25] ______, *The best explicatum for weight of evidence*, Journal of Statistical Computation and Simulation **19** (1984), 294–299.

[26] ______, *Weight of evidence: a brief survey*, Bayesian Statistics 2, North-Holland, 1985, pp. 249–269.

[27] ______, *A Reinstatement, in Response to Gillies, of Redhead's Argument in Support of Induction*, Philosophy of Science **54** (1987), 470–72.

[28] A. Hájek, *What conditional probabilities could not be*, *Synthese* (forthcoming), 2002.

[29] D. Heckerman, *An axiomatic framework for belief updates*, Uncertainty in Artificial Intelligence 2, Elsevier, 1988, pp. 11–22.

[30] E. Horvitz and D. Heckerman, *The inconsistent use of certainty measures in artificial intelligence research*, Uncertainty in Artificial Intelligence 1, Elsevier, 1986, pp. 137–151.

[31] P. Horwich, *Probability and evidence*, Cambridge University Press, 1982.

[32] C. Howson and P. Urbach, *Scientific reasoning: The bayesian approach*, Open Court, 1993.

[33] P.J. Hurley, *A concise introduction to logic*, eighth ed., Wadsworth, 2003.

[34] R. Jeffrey, *Probability and the art of judgment*, Cambridge University Press, 1992.

[35] J. Joyce, *The foundations of causal decision theory*, Cambridge University Press, 1999.

[36] J. Kemeny and P. Oppenheim, *Degrees of factual support*, Phil. of Sci. **19** (1952), 307–324.

[37] J. Keynes, *A treatise on probability*, Macmillan, 1921.

[38] A. Kolmogorov, *Foundations of probability*, second english ed., AMS Chelsea Publishing, 1956.

[39] H. Kyburg, *Recent work in inductive logic*, Recent Work in Philosophy (T. Machan and K. Lucey, eds.), Rowman & Allanheld, 1983, pp. 87–150.

[40] J. Mackie, *The relevance criterion of confirmation*, The British Journal for the Philosophy of Science **20** (1969), 27–40.

[41] P. Maher, *Probabilities for two properties*, Erkenntnis **52** (2000), no. 1, 63–91.

[42] ______, *Probabilities for multiple properties: the models of Hesse and Carnap and Kemeny*, Erkenntnis **55** (2001), no. 2, 183–216.

[43] P. Milne, $\log[p(h/eb)/p(h/b)]$ *is the one true measure of confirmation*, Philosophy of Science **63** (1996), 21–26.

[44] H. Mortimer, *The logic of induction*, Prentice Hall, 1988.

[45] R. Nozick, *Philosophical explanations*, Harvard University Press, Cambridge, 1981.

[46] J. Pearl, *Probabilistic reasoning in intelligent systems: Networks of plausible inference*, Morgan Kauffman, San Francisco, 1988.

[47] C. Peirce, *The probability of induction*, Popular Science Monthly **12** (1878), 705–718.

[48] K. Popper, *Degree of confirmation*, The British Journal for the Phil. of Sci. **5** (1954), 143–149.

[49] ______, *The logic of scientific discovery*, Routledge, 1992.

[50] K. Popper and D. Miller, *The impossibility of inductive probability*, Nature **302** (1983), 687–688.

[51] M. Redhead, *On the impossibility of inductive probability*, The British Journal for the Philosophy of Science **36** (1985), 185–191.

[52] H. Reichenbach, *The direction of time*, University of California, 1956.

[53] P. Roeper and H. Leblanc, *Probability theory and probability logic*, University of Toronto, 1999.

[54] R. Rosenkrantz, *Foundations and applications of inductive probability*, Ridgeview, 1981.

[55] ______, *Bayesian confirmation: Paradise regained*, The British Journal for the Philosophy of Science **45** (1994), 467–476.

[56] D. Schum, *The evidential foundations of probabilistic reasoning*, John Wiley & Sons, 1994.

[57] B. Skyrms, *Choice and chance*, fourth ed., Wadsworth, 2000.

[58] E. Sober, *Independent evidence about a common cause*, Philosophy of Science **56** (1989), 275–287.

[59] ______, *No model, no inference: A bayesian primer on the grue problem*, Grue! The New Riddle of Induction (D. Stalker, ed.), Open Court, 1994.

[60] L. Wittgenstein, *Philosophical investigations*, Macmillan, 1953.