STUDIES IN BAYESIAN CONFIRMATION THEORY

By

Branden Fitelson

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
(PHILOSOPHY)

at the

UNIVERSITY OF WISCONSIN – MADISON

2001

For Tina

The only good is knowledge and
the only evil is ignorance.
— Socrates

Abstract

According to Bayesian confirmation theory, evidence E (incrementally) confirms (or supports) a hypothesis H (roughly) just in case E and H are positively probabilistically correlated (under an appropriate probability function Pr). There are many logically equivalent ways of saying that E and H are correlated under Pr. Surprisingly, this leads to a plethora of *non*-equivalent quantitative measures of the *degree* to which E confirms H (under Pr). In fact, many non-equivalent Bayesian measures of the degree to which E confirms (or supports) H have been proposed and defended in the literature on inductive logic. I provide a thorough historical survey of the various proposals, and a detailed discussion of the philosophical ramifications of the differences between them. I argue that the set of candidate measures can be narrowed drastically by just a few intuitive and simple desiderata. In the end, I provide some novel and compelling reasons to think that the correct measure of degree of evidential support (within a Bayesian framework) is the (log) likelihood ratio. The central analyses of this research have had some useful and interesting byproducts, including: (*i*) a new Bayesian account of (confirmationally) independent evidence, which has applications to several important problems in confirmation theory, including the problem of the (confirmational) value of evidential diversity, and (*ii*) novel resolutions of several problems in Bayesian confirmation theory, motivated by the use of the (log) likelihood ratio measure, including a reply to the Popper-Miller critique of probabilistic induction, and a new analysis and resolution of the problem of irrelevant conjunction (*a.k.a.*, the tacking problem).

Acknowledgements

Without friends no one would choose to live,
despite having all other goods.
— Aristotle

I have many people to thank, and I apologize in advance if I have forgotten anyone. First, and foremost, I must thank my wife, Tina Eliassi–Rad, for her constant love, support, and encouragement throughout my graduate career, which has imbued everything in my life with value. I would also like to thank all of my friends and colleagues whose collaborative and personal support was essential for the completion of this research. In particular, I must first thank my advisor, Malcolm Forster, for convincing me to become a philosophy graduate student and for always making sure that I received the kind of support that I needed along the way. Special thanks also go out to Elliott Sober for taking such a personal interest in my career, and for being such a great collaborator and friend over the past several years. Ellery Eells and Patrick Maher deserve thanks for teaching me much of what I know about Bayesian philosophy of science. Indeed, the ideas for this project were first sparked by some immature gropings of mine in a seminar of Ellery Eells' on confirmation theory, which I attended (as an undergraduate) in 1992. The project really began to take shape in the next few years, after many discussions with Malcolm Forster. Finally, while I was studying Bayesian philosophy of science as a visiting scholar at UIUC under Patrick Maher in 1995, the thesis project (in its final form) was born. Among non-philosophers-of-science, Terry Penner and Larry Wos have been my most ardent supporters. Many others have contributed in various ways to my wonderful 'career' as a graduate student. Some of these are (in alphabetical order): Jim Anderson, Marty Barrett, Mike Byrd, Zac Ernst, Shelley Glodowski, Lori Grant, Alan Hájek, Ken Harris, Dan Hausman, Jim Jennings, Jim Joyce, Mike Kruse, Greg Mougin, Steve Schmid, Larry Shapiro, Jude Shavlik, Brian Skyrms, Doug Smith, Dan Steel, Chris Stephens, and Patty Winspur. Last, but not least, I must thank my parents for instilling in me both curiosity and determination (useful attributes for someone pursuing a PhD).

List of Figures

List of Tables

Contents

Chapter 1

Introduction

A good notation has a subtlety and suggestiveness
which at times make it almost seem like a live teacher.
— Bertrand Russell

1.1 Notation

Before we get started, some explanation of our basic notation is required. Table 1, below, should do the trick, for now (some special, additional notation will be introduced, as the monograph unfolds). With these basic notational conventions out of the way, we're ready for a brief introduction to Bayesian confirmation theory.

1.2 Background on Bayesian Confirmation

1.2.1 Probability Theory I: Kolmogorov's Axioms

The basic conceptual building block underlying Bayesian confirmation theory is the mathematical theory of probability. For our purposes, we won't need all of probability theory, just (a simple fragment of) the probability calculus. I will assume, throughout, that Pr is a Kolmogorov probability function. That is, I will assume that Pr (taken to be defined over a Boolean algebra of *propositions*) satisfies (for all propositions X and Y) the following axioms [Kolmogorov (1956)]:

$$\Pr(X) \geq 0.$$

$$\Pr(\top) = 1.$$

$$\text{If } X \,\&\, Y \vDash \bot, \text{ then } \Pr(X \vee Y) = \Pr(X) + \Pr(Y).$$ [1]

X, Y, H_1, E', ...	propositions
$(\forall x)$	universal quantifier
$(\exists x)$	existential quantifier
$\&$	conjunction (object language), $X \,\&\, Y$ (X and Y)
$\vee$	disjunction (object language), $X \vee Y$ (X or Y)
$\bar{X}$	negation (object language), $\bar{X}$ (not X)
$\top$	the necessary proposition
$\bot$	the impossible proposition
$=_{df}$	definition (metalanguage)
iff, $\Longleftrightarrow$	if and only if (metalanguage)
$\Longrightarrow$	only if (metalanguage)
$\models X$ (or $X = \top$)	X is logically true (*i.e.*, X is true in every possible world or model)
$X \models Y$	X entails Y (*i.e.*, Y is true in every possible world or model in which X is true)
$\log(x)$	the logarithm (base > 1) of (a real number) x
$\Pr(X)$	the (unconditional) probability of X, under Pr
$\Pr(X \mid Y)$	the probability of X, conditional on Y, under Pr

Table 1: The basic notation used in this monograph.

Moreover, following Kolmogorov, we will take the *conditional* probability $\Pr(\cdot \mid \cdot)$ to be *defined* in terms of the *un*conditional probability $\Pr(\cdot)$, as follows.[2]

$$\Pr(X \mid Y) =_{df} \frac{\Pr(X \,\&\, Y)}{\Pr(Y)} \qquad [\text{where } \Pr(Y) \neq 0]$$

[1]Sometimes, Kolmogorov's third axiom is strengthened to require additivity over *countably infinite* sets of mutually exclusive propositions. This stronger assumption (called *countable additivity*) will *not* be made here. Typically, my discussion will require only *finite* spaces.

[2]Some have argued [most forcefully, Hájek (2001)] that we should, instead, take conditional probabilities as *primitive*, and then define unconditional probabilities in terms of conditional probabilities [as Popper (1980, Appendix *iv) and others have]. While I am somewhat sympathetic to this suggestion (in some contexts), I have chosen *not* to make this move here. Doing so would require making the arguments below far more complicated, and far less intuitive (see footnote 10). As far as I know, only one philosopher of science has commented (in print) on the ramifications of such a move for Bayesian confirmation theory [see Festa (1999)].

1.2.2 Probability Theory II: Interpretation(s)

Happily, almost all of the arguments and analyses below will *not* trade on the *interpretation* of the probability function Pr. This is good news, since there is great and well-known controversy over the interpretation, origin, and status of unconditional (or "*a priori*") probabilities $\Pr(\cdot)$. As I mentioned above, all of the probabilities in this monograph are defined in terms of unconditional probabilities. So, if these cannot be well understood or well grounded in a Bayesian framework, then it's hard to see how Bayesian confirmation could even get off the ground.[3] The issues raised and the arguments constructed below will retain their interestingness and legitimacy under objective, logical, or epistemic interpretations of probability.[4] So, I will not argue for the existence of a 'rational' (or 'objective') probability function Pr of the kind required to give Bayesian confirmation theory its (objective) normative teeth.[5] Hereafter, I will simply assume that the problems of identifying and interpreting a suitable Pr have been solved. I will argue below that even if (*per impossible*?) all of these foundational problems involving the status and interpretation of Pr in the Bayesian framework were solved, interesting problems would remain in Bayesian confirmation theory. Now, we're ready to introduce the basic concepts to be investigated for the remainder of the monograph.[6]

[3]See Earman (1992) for a thorough discussion (including many historical references) of the problem of priors and other foundational controversies in Bayesian philosophy of science.

[4]See Fine (1973) and Gillies (2000) for nice surveys of the wide variety of interpretations of probability (both subjective and objective) that have been propounded over the years.

[5]Maher (1996) discusses the subjective *versus* the objective in Bayesian confirmation theory.

[6]The following books and articles on various issues surrounding the role of probability in philosophy, science, and statistics (though not explicitly cited elsewhere in this monograph) have had a significant influence on my views: Adams (1998), Carnap (1971), Dale (1999), Deák (1990), de Finetti (1990), Eells (1983, 1985), Efron (1978, 1986), Feller (1968), Forster (1994, 2000), Forster and Kieseppä (2001), Forster and Sober (1994), Glymour (1980), Hacking (1975), Hailperin (1996), Hempel (1945, 1983), Jeffreys (1998), Levi (1967), Maher (1993, 2000, 2001), Maĭstrov (1974), Paris (1994), Ramsey (1990), Reichenbach (1971), Robert (1994), Rosenkrantz (1977), Royall (1997), Savage (1972), Schervish (1995), Seidenfeld (1979), Skyrms (1984), Sober (1994a, 1994c), Stigler (1990), Swinburne (1973), Székely (1986), van Fraassen (1982, and 1989).

1.2.3 Qualitative Bayesian Confirmation

As Carnap (1962, new preface) was one of the first to make clear, there are two distinct kinds of Bayesian confirmation. That is to say, there are two ways of understanding the relation "E confirms H, relative to background knowledge[7] K" within a Bayesian framework. Bayesians might use either of the following:

- E confirms H relative to K iff $\Pr(H \mid E \,\&\, K) > k$ [for some $k \in (0, 1)$].
- E confirms H relative to K iff $\Pr(H \mid E \,\&\, K) > \Pr(H \mid K)$.

The first of these is what Carnap calls "confirmation as firmness" and the second is what Carnap calls "confirmation as increase in firmness." This monograph is concerned *only* with the latter Bayesian notion of confirmation. In the contemporary literature, this is known as *incremental* (Bayesian) confirmation.[8] This leads us to our first definition: that of the of the qualitative, ternary relation "E (incrementally) confirms H relative to background K."

> **Definition**. E *confirms* H relative to K iff $\Pr(H \mid E \& K) > \Pr(H \mid K)$. If $\Pr(H \mid E \,\&\, K) < \Pr(H \mid K)$, then we say that E *dis*confirms H relative to K. And if $\Pr(H \mid E \,\&\, K) = \Pr(H \mid K)$, then we say that E *is confirmationally irrelevant to* H relative to K.

The remainder of our discussion will be concerned with various quantitative generalizations of this qualitative, incremental variety of Bayesian confirmation.

[7]The word "knowledge" prejudices us here toward an *epistemic* reading of Pr. If you prefer, you may think of K (instead) simply as a set of background propositions, which may or may not be known by any agent. As I mentioned above, none of my arguments will trade on this distinction between objective and subjective probability (or confirmation). I prefer to talk in terms of epistemic probabilities in this context, but the reader may have a different preference.

[8]The terms "absolute" and "incremental" are used nowadays to denote the two varieties of Bayesian confirmation discussed in Carnap (1962, new preface). It seems that incremental confirmation is used almost exclusively by contemporary Bayesians. See Earman (1992).

1.2.4 Quantitative Confirmation I: The Basic Concepts

If one adopts Kolmogorov's theory of probability, then there are many different (but *logically equivalent*) ways of saying that E confirms H relative to K. This is because, according to our definition, E confirms H relative to K iff *E and H are positively correlated under* $\Pr(\cdot \mid K)$.[9] It is well known that the following are four (logically equivalent) ways of saying that E and H are positively correlated under $\Pr(\cdot \mid K)$, provided that $\Pr(\cdot \mid K)$ is a Kolmogorov probability function.[10]

- $\Pr(H \mid E \,\&\, K) > \Pr(H \mid K)$
- $\Pr(H \mid E \,\&\, K) > \Pr(H \mid \bar{E} \,\&\, K)$
- $\Pr(E \mid H \,\&\, K) > \Pr(E \mid \bar{H} \,\&\, K)$
- $\Pr(H \,\&\, E \mid K) > \Pr(H \mid K) \cdot \Pr(E \mid K)$

Intuitively, one might propose to define various *quantitative* measures of the *degree* to which E confirms H relative to K, using the inequalities above (or any other equivalent inequality). For instance, by taking the difference between the left and right hand side of any of these inequalities, one may construct a (intuitively plausible) measure $\mathfrak{c}(H, E \mid K)$ of the degree to which E confirms H relative to K. Any such measure is bound to satisfy the following *qualitative* constraint, in cases where E confirms, disconfirms, or is confirmationally irrelevant to H, given K:

[9] Correlation and dependence of random variables do *not* coincide — *in general*. However, in the special case (at hand) of *dichotomous* random variables (*i.e.*, *propositions*), correlation and dependence are synonymous. See Ross (1994, pages 320–331) for discussion.

[10] Interestingly (because of the fundamentally different way in which such theories handle conditional probabilities with impossible antecedents), this claim is *not* true for theories of probability that take conditional probabilities as primitive [*e.g.*, those of Popper (1980, Appendix *iv), Carnap (1962, §53), and others]. This explains why the arguments below would require serious revision if these alternative ways of axiomatizing Pr were used. See Roeper and Leblanc (1999) and Goosens (1979) for discussions of the various alternative ways of defining probabilities by taking conditional probabilities as primitive. And, see Festa (1999) for a taste of the disunifying effects this would have on both qualitative and quantitative Bayesian confirmation theory.

$$(\mathcal{R}) \qquad \mathfrak{c}(H, E \mid K) \begin{cases} > 0 & \text{if } \Pr(H \mid E \,\&\, K) > \Pr(H \mid K), \\ < 0 & \text{if } \Pr(H \mid E \,\&\, K) < \Pr(H \mid K), \\ = 0 & \text{if } \Pr(H \mid E \,\&\, K) = \Pr(H \mid K). \end{cases}$$

Any measure $\mathfrak{c}(H, E \mid K)$ that satisfies $\mathcal{R}$ will be called a *relevance measure*.[11] One can also form relevance measures by taking logarithms of ratios of the left and right hand sides of any of the inequalities above. Because any two relevance measures must be *qualitatively* equivalent (*viz.*, $\mathcal{R}$), one might suspect that it really shouldn't matter too much which relevance measure one uses. While different relevance measures will assign different *numerical* values, one might expect them to impose the same *ordering* over H's, E's and K's. In other words, one might expect all relevance measures to be *ordinally equivalent* in the following precise sense.[12]

> **Definition**. Two measures $\mathfrak{c}_1(H, E \mid K)$ and $\mathfrak{c}_2(H, E \mid K)$ of the degree to which E confirms H relative to K are said to be *ordinally equivalent* just in case, for all H, E, K, H', E', and K':
>
> $$\mathfrak{c}_1(H, E \mid K) \geq \mathfrak{c}_1(H', E' \mid K') \text{ iff } \mathfrak{c}_2(H, E \mid K) \geq \mathfrak{c}_2(H', E' \mid K').$$

The surprising thing (and *the* central fact in this monograph) is that this intuition couldn't be farther from the truth. It is rather surprising (and it has not been widely noticed or discussed), but most proposed relevance measures — although stemming from the very same *qualitative* notion — give rise to (radically) *non*-equivalent *quantitative* gauges of the *degree* to which E confirms H relative to K. This radical, ordinal disagreement between the many proposed relevance measures of degree of confirmation is the main issue that I will address below.

[11]Measures that *violate* $\mathcal{R}$ (*i.e.*, *non*-relevance-measures) will not be considered measures of *confirmation* in the sense defined above. That is, $\mathcal{R}$ will be taken as a minimal *desideratum* for any adequate measure $\mathfrak{c}(H, E \mid K)$ of the degree to which E confirms H relative to K.

[12]See Krantz, Luce, Suppes, and Tversky (1971, Ch. 1) for a theoretical treatment of the ordinal equivalence of abstract quantitative measures.

1.2.5 Quantitative Confirmation II: The Many Measures

As I mentioned above, taking differences or logarithms of ratios of the left and right hand sides of the relevant inequalities is an easy way to generate relevance measures. There are also lots of other, more complicated ways of generating relevance measures. In the history of inductive logic, a great many such measures have been proposed and defended.[13] Rather than give an exhaustive list of all of these measures (there are dozens), I will only discuss the following five relevance measures, which are representative of the kinds of measures that have been proposed and defended (especially, in recent years), and which cover much of the space of possible ordinal structures imposable by (simple) relevance measures.[14]

$$d(H, E \mid K) =_{df} \Pr(H \mid E \,\&\, K) - \Pr(H \mid K)$$

$$r(H, E \mid K) =_{df} \log\left[\frac{\Pr(H \mid E \,\&\, K)}{\Pr(H \mid K)}\right]$$

$$\begin{aligned} l(H, E \mid K) &=_{df} \log\left[\frac{\Pr(E \mid H \,\&\, K)}{\Pr(E \mid \bar{H} \,\&\, K)}\right] \\ &= \log\left[\frac{\Pr(H \mid E \,\&\, K) \cdot [1 - \Pr(H \mid K)]}{[1 - \Pr(H \mid E \,\&\, K)] \cdot \Pr(H \mid K)}\right]. \end{aligned}$$ [15]

[13]For a nice survey, see Kyburg (1983).

[14]Advocates of d include Earman (1992), Eells (1982), Gillies (1986), Jeffrey (1992), and Rosenkrantz (1994). Advocates of r (or measures ordinally equivalent to r) include Horwich (1982), Keynes (1921), Mackie (1969), Milne (1996), Schlesinger (1995), and Pollard (1999). Advocates of l (or measures ordinally equivalent to l) include Kemeny and Oppenheim (1952), Good (1983), Heckerman (1988), Horvitz and Heckerman (1986), Pearl (1988), and Schum (1994). Recent proponents of s include Christensen (1999) as well as Joyce (1999). $\mathfrak{r}$ is Carnap's (1962, §67) relevance measure. Logarithms (base > 1) of the ratios $\Pr(H \mid E \,\&\, K)/\Pr(H \mid K)$ and $\Pr(E \mid H \,\&\, K)/\Pr(E \mid \bar{H} \,\&\, K)$ are taken to insure that (i) r and l satisfy $\mathcal{R}$, and (ii) r and l are *additive* in various ways. Not all advocates of r or l adopt this convention (*e.g.,* Horwich (1982)). But, because logarithms are isotone, defining r and l in this way will not effect their ordinal structure. Hence, using logarithms will not effect the generality of our subsequent arguments.

$$s(H, E \mid K) =_{df} \Pr(H \mid E \,\&\, K) - \Pr(H \mid \bar{E} \,\&\, K)$$
$$= \frac{1}{\Pr(\bar{E} \mid K)} \cdot d(H, E \mid K).$$ [16]

$$\mathfrak{r}(H, E \mid K) =_{df} \Pr(H \,\&\, E \,\&\, K) \cdot \Pr(K) - \Pr(H \,\&\, K) \cdot \Pr(E \,\&\, K)$$
$$= \Pr(K) \cdot \Pr(E \,\&\, K) \cdot d(H, E \mid K).$$ [17]

In the remaining chapters, I will (aim to): (*i*) show that these five popular relevance measures disagree in some very important ways, (*ii*) discuss how these disagreements effect many existing philosophical arguments, and (*iii*) explain how the field of competing measures can be drastically narrowed using just a few simple, intuitive principles. Along the way, I will also provide some new analyses (and resolutions) of some old problems in Bayesian confirmation theory.

---

[15]Expressing $l(H, E \mid K)$ in this way makes it clear that, like $d(H, E \mid K)$ and $r(H, E \mid K)$, $l(H, E \mid K)$ is a function of the posterior $\Pr(H \mid E \,\&\, K)$ and prior $\Pr(H \mid K)$ of H (given K).

[16]This equality holds provided, of course, that $\Pr(\bar{E} \mid K) \neq 0$. See Christensen (1999) for further discussion about the relationship between d and s.

[17]It is perhaps easiest to think of Carnap's $\mathfrak{r}$ as a kind of *covariance* measure. Indeed, when K is *tautologous*, we have: $\mathfrak{r}(H, E \mid K) = \Pr(H \,\&\, E \,\&\, K) \cdot \Pr(K) - \Pr(H \,\&\, K) \cdot \Pr(E \,\&\, K) = \Pr(H \,\&\, E) - \Pr(H) \cdot \Pr(E) = \mathrm{Cov}(H, E)$. In general, $\mathfrak{r}(H, E \mid K) = \Pr(K)^2 \cdot \mathrm{Cov}(H, E \mid K)$.

Chapter 2

The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity[18]

Measure what is measurable, and
make measurable what is not so.
— Galileo

2.1 A General Overview of the Problem

Many arguments surrounding quantitative Bayesian confirmation theory presuppose that the degree to which E incrementally confirms H, given K should be measured using some relevance measure (or, class of relevance measures) $\mathfrak{c}$, where $\mathfrak{c}$ is taken to have a certain (ordinal) structure. We say that an argument $\mathcal{A}$ of this kind is *sensitive to choice of measure* if $\mathcal{A}$'s validity varies (*ceteris paribus*), depending on which of the five relevance measures d, r, l, s, or $\mathfrak{r}$ is used in $\mathcal{A}$. If $\mathcal{A}$ is valid *regardless* of which of the five relevance measures d, r, l, s, or $\mathfrak{r}$ is used in $\mathcal{A}$, then $\mathcal{A}$ is said to be *in*sensitive to choice of measure (or, simply, *robust*).[19]

Below, I will show that eight well-known and central arguments surrounding

[18] Much of the material in this chapter appears in Fitelson (1999).

[19] One can invent more or less stringent varieties of measure sensitivity. For instance, one could call an argument "measure sensitive" (in a *very strict* sense) if $\mathcal{A}$ is valid with respect to *some conceivable* relevance measure $\mathfrak{c}_1$, but invalid with respect to some other *conceivable* relevance measure $\mathfrak{c}_2$. Of course, such a strict notion of sensitivity would probably not be very interesting, since highly gerrymandered relevance measures can undoubtedly be concocted to suit arbitrary purposes. I am employing a much less strict notion of measure sensitivity which appeals only to measures that have actually been used and defended in the philosophical literature.

contemporary Bayesian confirmation theory are sensitive to choice of measure. I will argue that this exposes a weakness in the theoretical foundation of Bayesian confirmation theory which must be shored-up. I call this problem *the problem of measure sensitivity*. After presenting a survey of measure sensitive arguments, I will examine some recent attempts to resolve the measure sensitivity problem. I will argue that, while some progress has been made toward this end, we still do not have an adequate or a complete resolution of the measure sensitivity problem. Specifically, I will show that the many defenders of the difference measure d have failed to provide compelling reasons to prefer d over the alternative measures l, s, and $\mathfrak{r}$. Thus, a pervasive problem of measure sensitivity still remains for many modern advocates and practitioners of Bayesian confirmation theory.

2.2 Contemporary Examples of the Problem

It is not difficult to show that *no pair of the five measures* d*,* r*,* l*,* s*, and* $\mathfrak{r}$ *is ordinally equivalent*. That is, each of these five measures can impose *distinct orderings* over sets of hypotheses and collections of evidence.[20] I have not seen many discussions concerning the measure sensitivity of concrete arguments surrounding Bayesian confirmation theory.[21] In this section, I will show that a wide variety of well-known arguments surrounding Bayesian confirmation theory are sensitive to choice of measure. Indeed, *almost every* argument I have seen surrounding Bayesian confirmation theory (that is, every Bayesian argument involving quantitative confirmational comparisons) turns-out to be sensitive to choice of measure!

[20]Rosenkrantz (1981, Exercise 3.6) discusses the ordinal non-equivalence of d, r, and l. Carnap (1962, §67) talks about some important ordinal differences between $\mathfrak{r}$, d, and r (Carnap does not compare $\mathfrak{r}$ with l or s). Joyce (1999) and Christensen (1999) discuss some of the ordinal differences among s, d, r, and l. And, Good (1985) mentions some differences among l, d, and r. As far as I know, notwithstanding these piecemeal discussions (and a few others like them), there has not been a thorough treatment of the ordinal disagreements between relevance measures.

[21]Two notable exceptions are Redhead (1985) and Mortimer (1988, §11.1).

2.2.1 Gillies's Rendition of the Popper-Miller Argument

Gillies (1986) reconstructs the infamous argument of Popper and Miller (1983) for the "impossibility of inductive probability" in such a way that it trades essentially on the following *additivity* property of the difference measure d:

$$(1) \qquad d(H, E \mid K) = d(H \vee E, E \mid K) + d(H \vee \bar{E}, E \mid K).$$

The details of the Popper-Miller argument need not concern us. All that matters for our purposes is that Gillies's rendition of the Popper-Miller argument against Bayesianism presupposes that any adequate Bayesian measure of inductive support (or confirmation) will satisfy the additivity property depicted in (1).

Redhead (1985) points out that *not all* Bayesian relevance measures have this requisite additivity property. Specifically, Redhead (1985) notes that the log-ratio measure r does *not* satisfy (1). It follows that the Popper-Miller argument is *sensitive to choice of measure*. Gillies (1986) responds to Redhead's point by showing that the measure r is *not* an adequate Bayesian relevance measure of confirmation. Gillies argues that the ratio measure r is inferior to the difference measure d because r fails to cope properly with cases of *deductive evidence* (see §2.3.1 for more on this argument against r). Unfortunately, Gillies fails to recognize that Redhead's criticism of the Popper-Miller argument can be significantly strengthened *via* the following theorem (see the Appendix for proofs of all Theorems).[22]

Theorem 1. *l does not have the additivity property expressed in* (1).[23]

Moreover, as we will see below in §2.3.1, the log-likelihood ratio measure l is immune to Gillies's criticism of r. So, pending some good reason to prefer d over

[22]This point was made, independently, by Good (1987).

[23]Carnap's relevance measure $\mathfrak{r}$ *does* satisfy (1). This follows easily from (1), and the fact that $\mathfrak{r}(H, E \mid K) = \Pr(K) \cdot \Pr(E \,\&\, K) \cdot d(H, E \mid K)$. Christensen's measure s also satisfies (1). This is an easy consequence of (1) and the fact that $s(H, E \mid K) = \dfrac{d(H, E \mid K)}{\Pr(\bar{E} \mid K)}$ (proofs omitted).

l, Gillies's reconstruction of the Popper-Miller argument does not seem to pose a serious threat to Bayesian confirmation theory (charitably reconstructed). What makes this first example so interesting is that it illustrates that the problem of measure sensitivity effects not only arguments *within* the Bayesian framework, but also arguments that are *critical of* Bayesianism.

2.2.2 Rosenkrantz *&* Earman on "Irrelevant Conjunction"

Rosenkrantz (1994) offers 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 d:

(2) If $H \vDash E$, then $d(H \,\&\, X, E \mid K) = \Pr(X \mid H \,\&\, K) \cdot d(H, E \mid K)$.

I won't bother to get into the details of Rosenkrantz's argument here (see §2.2.2.1 below for a detailed analysis). It suffices, for my present purposes, to note that it depends sensitively on property (2). As a result, Rosenkrantz's argument does *not* go through if one uses r or l, instead of d, to measure degree of confirmation. The proof of the following theorem demonstrates this strong measure sensitivity of Rosenkrantz's approach:

Theorem 2. *Neither r nor l has the property expressed in* (2).[29]

Consequently, Rosenkrantz's account of "irrelevant conjunction" is adequate only if the difference measure d is to be preferred over the two alternative relevance measures r and l. Like Gillies, Rosenkrantz (1981, Exercise 3.6) does provide some good reasons to prefer d over r (see §2.3.1). However, he explicitly admits that he knows of "no compelling considerations that adjudicate between" the difference measure d and the log-likelihood ratio measure l. This leaves Rosenkrantz in a rather uncomfortable position. As I will discuss below, Rosenkrantz is not alone in this respect. I know of no arguments (much less, compelling ones) that have been proposed to demonstrate that d should be preferred over l.

Earman (1992) offers a similar approach to "irrelevant conjunction" which is less sensitive to choice of measure. Earman's approach relies only on the following logically weaker fact about d:

$$(2') \qquad \text{If } H \vDash E\text{, then } d(H \,\&\, X, E \mid K) < d(H, E \mid K).$$

The log-likelihood ratio measure l, Carnap's measure $\mathfrak{r}$, and Christensen's measure s all satisfy (2′) (proofs omitted); but, the log-ratio measure r does *not* satisfy (2′) (see §2.3.1). So, while still sensitive to choice of measure, Earman's "irrelevant conjunction" argument is *less* sensitive to choice of measure than Rosenkrantz's.

2.2.2.1 A New Analysis of "Irrelevant Conjunction"[24]

The problem of irrelevant conjunction (*a.k.a.*, the tacking problem) was originally raised as a problem for the hypothetico-deductive (H–D) account of confirmation.[25] According to the H–D account of confirmation, E confirms H relative to K if (roughly) $H \,\&\, K \vDash E$. Therefore, owing to the monotonicity of $\vDash$, we have the following fact about H–D-confirmation:

(3) If E H–D-confirms H relative to K, then E also H–D-confirms $H \& X$ relative to K, for *any* X.

The problem with (3) is supposed to be that conjoining X's that are *utterly irrelevant* to H and E seems (intuitively) to undermine the confirmation E provides

[24] The material for this section [not appearing in Fitelson (1999)] is taken from Fitelson (2001b).

[25] See Hempel (1945) for the original (classic) presentation of H-D confirmation, and some of its shortcomings (including the problem of irrelevant conjunction). See Skyrms (1992) for an incisive and illuminating critical survey of some recent papers on deductive accounts of confirmation and the problem of irrelevant conjunction. And, see Earman (1992, pp. 63–65) for a typical Bayesian discussion of the problem of irrelevant conjunction.

for the resulting (conjunctive) hypothesis H & X. For instance, intuitively, the return of Halley's comet in 1758 (E) confirmed Newton's theory (H) of universal gravitation (relative to the background evidence (K) available at the time). But, according to the H–D account of confirmation, this implies that the return of Halley's comet also confirms the conjunction of H and (say) Coulomb's Law (or any other proposition(s) one would like to conjoin to H). And, no matter how many irrelevancies are conjoined to H, E will continue to confirm the conjunction, according to the H–D account of confirmation.

Because probabilistic correlation is *not* monotonic, Bayesian confirmation does *not* have the property expressed in (3). That is, according to Bayesianism, it does *not* follow from E's confirming H that E must also confirm H & X, for any X. So, Bayesianism is immune from the original problem of irrelevant conjunction. However, Bayesian confirmation does still suffer from this problem in the case of *deductive evidence* (*i.e.*, in the H–D case in which $H \,\&\, K \vDash E$). That is, Bayesian confirmation and H–D confirmation both satisfy the following *special case* of (3):

(3′) If $H \,\&\, K \vDash E$, then E confirms H & X relative to K, for *any* X.

Bayesians (*e.g.*, Rosenkrantz and Earman, as discussed above) have tried to resolve this *new* problem of irrelevant conjunction by proving various *quantitative* results about the *degree* to which E confirms H *versus* H & X in the case of deductive evidence. As I have discussed above, theorems like Earman's (2′) are typically called into action in this context. Bayesians will explain that, while E does continue to confirm $H \,\&\, X$ in the case of deductive evidence, the *degree* to which E confirms $H \,\&\, X$ will be *less than* the degree to which E confirms H. And, as more and more irrelevant conjuncts are added, the degree to which E confirms the conjunction will tend to decrease. I have already shown that this claim is *sensitive to choice of measure* (since it is not true for measure r). But, there is an even more serious philosophical flaw in the standard Bayesian analyses of this problem.

A closer look at (2′) reveals that the *irrelevance* of X has *disappeared* from the Bayesian resolution of the problem of *irrelevant* conjunction. What (2′) says is that, as *conjuncts* X (*simpliciter*) are added to H, the degree to which E confirms H & X will tend to decrease. As far as (2′) is concerned, a conjunct X could be (intuitively) *relevant* to H and E, but this would not prevent the conjunction $H\&X$ from being less strongly confirmed than H by E. This is unfortunate, for two reasons. First, it was supposed to be the *irrelevant* X's that made (3) and (3′) seem so unattractive. It's not so obvious that either (3) or (3′) is incorrect in the case of *highly positively relevant* X's. Moreover, Bayesian confirmation theory is founded on a perfectly precise and intuitive kind of *relevance* (*viz.*, *correlation*), which is not mentioned anywhere in either the ("Bayesian") statement(s) or resolution(s) of the problem of irrelevant conjunction. So, Bayesians who endorse Earman's resolution [grounded in (2′)] have apparently both (*i*) lost track of *which* X's were supposed to make (3) and (3′) seem so unintuitive; and, in the process, (*ii*) forsaken the very notion of *relevance* that undergirds their own theory. Rosenkrantz (1994, pp. 470–471) does seem somewhat sensitive to these points. He motivates his resolution (*viz.*, (2)) of the problem of irrelevant conjunction, as follows.[26]

> On H–D accounts, H is confirmed by a verified prediction, E, but E is equally a prediction of H & X, where the 'tacked on' X may be a quite extraneous hypothesis. ... There are those who think that this sin of 'irrelevant conjunction' vitiates Bayesian confirmation theory as well. ... I hope you will agree that the two extreme positions on this issue are equally unpalatable, (i) that a consequence E of H confirms H not at all, and (ii) that E confirms H & X just as strongly as it confirms H alone. ... In general, intuition expects intermediate degrees

[26] I have translated Rosenkrantz's (1994, pp. 470–471) passage into our notation.

> of confirmation that depend on the degree of compatibility of H with X. Measuring degree of confirmation by ... [d] ... yields ... [(2)].

Rosenkrantz deserves credit here for *trying* to bring the *irrelevance* back into the Bayesian resolution of this problem. However, his account has two serious flaws. First, as I have already shown, his account is more sensitive to choice of measure of confirmation than the generic Bayesian resolutions (*e.g.*, Earman's (2′)-based account): it only works if we adopt d, s, or $\mathfrak{r}$ as our measure of confirmation. In addition, Rosenkrantz's account makes use of a strange — and decidedly *non*-Bayesian — notion of "relevance." Rosenkrantz seems to be suggesting that a conjunct X should be considered "irrelevant" to H (relative to background K) if $\Pr(X \mid H \,\&\, K) < 1$.[27] This suggestion is inadequate for two reasons. First, since when do Bayesians think that the degree to which X is relevant to H can be measured using only the conditional probability $\Pr(X \mid H \,\&\, K)$? Secondly, the inequality $\Pr(X \mid H \,\&\, K) < 1$ can, at best, only tell us when X is "irrelevant" to H — it can say nothing about whether X is "irrelevant" to E, or to various logical combinations of H and E. It seems to me that the cases in which (3) and (3′) are *least* intuitive are cases in which X is (intuitively) irrelevant to both H and E, and to all logical combinations of H and E. We need a different approach here.

I suggest that we go about this in an entirely different way. Let's *start* by saying what it *means* (in a Bayesian framework) to say that X is confirmationally irrelevant to H, E, and all logical combinations H and E. Then, once we have this precisely defined, let's see if (and under what auxiliary assumptions) we can show that such *irrelevant* conjuncts lead to decreased confirmational power. The first of these tasks is already done. Bayesians *already* have a perfectly precise definition of confirmational irrelevance: probabilistic independence. Therefore, we already know what it means (in a Bayesian confirmation-theoretic framework) to say that

[27] This is the (necessary and sufficient) condition under which Rosenkrantz's (2) entails a decrease in the degree of confirmation E provides for $H \,\&\, X$ *versus* H (relative to K).

X is confirmationally irrelevant to H, E, and all logical combinations H and E. Finally, our story has a happy ending, in the form of the following general and robust theorem (here, "confirms" is used in the Bayesian, relevance sense):

Theorem 3. *If E confirms H, and X is confirmationally irrelevant to H, E, and $H \,\&\, E$ (relative to background K), then $\mathfrak{c}(H, E \mid K) > \mathfrak{c}(H \,\&\, X, E \mid K)$, where $\mathfrak{c}$ may be any of our five relevance measures, except r.*[28]

Our Bayesian resolution of the (new) problem of irrelevant conjunction has the following advantages over its existing rivals:

- Our resolution *makes use of the irrelevance of X*. Moreover, our notion of confirmational irrelevance is not some peculiar one (like Rosenkrantz's), but just the standard Bayesian concept, based on probabilistic independence.

- Our resolution is not restricted to the (not very inductively interesting) special case of *deductive evidence*; it explains why *irrelevant* conjuncts are confirmationally disadvantageous, in *all* contexts (deductive or otherwise).

- Our resolution is as robust as any other existing resolution (*e.g.*, Earman's), and more robust than any other existing account that tries to be sensitive to "irrelevance" of the conjunct X in some sense or other (*e.g.*, Rosenkrantz's).

2.2.3 Eells & Sober on the Grue Paradox

Eells (1982) offers a resolution of the Grue Paradox which trades on the following property of the difference measure d [where $\beta =_{df} \Pr(H_1 \,\&\, E \mid K) - \Pr(H_2 \,\&\, E \mid K)$,

[28] Apparently, r cannot be used to resolve the problem of irrelevant conjunction — even in cases that do not involve deductive evidence. This shows an even deeper problem with r than the (mere) "deductive insensitivity," which prevents r from satisfying Earman's ($2'$) (see §2.3.1). Defenders of r [*e.g.*, Milne (1996)] are quick to point out that r's (mis)handling of the traditional, *deductive* problem of irrelevant conjunction is not such a serious weakness of r. Indeed, Milne (1996) characterizes the traditional problem of irrelevant conjunction as a "wretched shibboleth." I am somewhat sympathetic to this point of view. *De*ductive cases are not terribly interesting from an *in*ductive-logical point of view. However, I think that r's mishandling of irrelevant conjuncts in the *inductive* case (as in Theorem 3) ought to be taken seriously. See footnote 36.

and $\delta =_{df} \Pr(H_1 \mathbin{\&} \bar{E} \mid K) - \Pr(H_2 \mathbin{\&} \bar{E} \mid K)]$:

$$(4) \qquad \text{If } \beta \cdot \frac{\Pr(\bar{E} \mid K)}{\Pr(E \mid K)} > \delta \text{ and } \Pr(E \mid K) < \frac{1}{2}, \text{ then } d(H_1, E \mid K) > d(H_2, E \mid K).$$

As usual, I will skip over the details of Eells's proposed resolution of Goodman's "new riddle of induction." What is important for our purposes is that (4) is *not* a property of either the log-likelihood ratio measure l or the log-ratio measure r, as is illustrated by the proof of the following theorem:

Theorem 4. *Neither r nor l has the property expressed in* (4).[29]

As a result, Eells's resolution of the Grue Paradox only works if one assumes that the difference measure d is to be preferred over the log-likelihood ratio measure l and the log-ratio measure r. Eells (personal communication) has described a possible reason to prefer d over r (this argument is discussed in §2.3.2). As far as I know, Eells has offered no argument to the effect that d should be preferred to l.

Sober (1994b) offers a similar approach to "Grue" which is less sensitive to choice of measure. Sober's approach relies only on the following logically weaker property of d:

$$(4') \qquad \text{If } H_1 \vDash E,\ H_2 \vDash E, \text{ and } \Pr(H_1 \mid K) > \Pr(H_2 \mid K), \text{ then } d(H_1, E \mid K) > d(H_2, E \mid K).$$

The log-likelihood ratio measure l, Carnap's relevance measure $\mathfrak{r}$, and Christensen's measure s all satisfy (4′) (proofs omitted); but, the log-ratio measure r does *not* satisfy (4′) (see §2.3.1). So, while still sensitive to choice of measure, Sober's "Grue" argument is *less* sensitive to choice of measure than Eells's.

[29] It is not difficult to show that (2) and (4) *do* hold for both $\mathfrak{r}$ and s (proofs omitted).

2.2.4 Horwich *et al.* on Ravens and Variety

A great many contemporary Bayesian confirmation theorists (including Horwich (1982)) have offered quantitative resolutions of the Ravens paradox *and/or* the problem of varied (or diverse) evidence which trade on the following relationship between conditional probabilities and relevance measures of confirmation.[30]

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

As it turns out (fortuitously), all three of the most popular contemporary relevance measures d, r, and l share property (5) (proofs omitted). But, neither Carnap's $\mathfrak{r}$ nor Christensen's s satisfies (5), as the proof of Theorem 5 shows.

Theorem 5. *Neither* $\mathfrak{r}$ *nor* s *has the property expressed in* (5).[31]

Until we are given some compelling reason to prefer d, r, and l to Carnap's $\mathfrak{r}$ and Christensen's s (and, to any other relevance measures which violate (5) — see footnote 31 and Appendix §A.5 for further discussion), we should be wary about accepting the popular quantitative resolutions of the Ravens Paradox, or the recent Bayesian accounts of the confirmational significance of evidential diversity.[32]

[30] An early quantitative resolution of the Ravens Paradox was given by Hosiasson-Lindenbaum (1940). Hosiasson-Lindenbaum was *not* working within a relevance framework. So, for her, it *was* sufficient to establish that $\Pr(H \mid E_1 \,\&\, K) > \Pr(H \mid E_2 \,\&\, K)$, where E_1 is a black-raven, E_2 is a non-black non-raven, H is the hypothesis that all ravens are black, and K is our background knowledge. Contemporary Bayesian relevance theorists have presupposed that this inequality is sufficient to establish that a black raven *incrementally* confirms that all ravens are black more strongly than a non-black non-raven does. As Theorem 5 shows, this is true for only *some* relevance measures. This same presupposition is also made by Bayesians who argue that (ceteris paribus) more varied sets of evidence (E_1) confirm hypotheses (H) more strongly than less varied sets of evidence (E_2) do. See Earman (1992, 69–79) for a survey of recent Bayesian resolutions of the Ravens Paradox, and Wayne (1995) for a survey of recent Bayesian resolutions of the problem of evidential diversity. As far as I know, *all* of these popular contemporary approaches are measure sensitive in the sense described here.

[31] There are other relevance measures which violate (5). Mortimer (1988, §11.1) shows that the measure $\Pr(E \mid H \,\&\, K) - \Pr(E \mid K)$ violates (5). It also turns out that Nozick's (1981, 252) measure $\Pr(E \mid H \,\&\, K) - \Pr(E \mid \bar{H} \,\&\, K)$ violates (5). See Appendix §A.5 for proofs.

[32] See Fitelson (1996) (and §3.3.2 below) and Wayne (1995) for independent reasons to be wary

2.2.5 An Important Theme in Our Examples

As our examples illustrate, several recent Bayesian confirmation theorists have presupposed the superiority of the difference measure d over one or more of the four alternative relevance measures r, l, s, and $\mathfrak{r}$. Moreover, we have seen that many well-known arguments in Bayesian confirmation theory depend sensitively on this assumption of d's superiority. To be sure, there are other arguments that fit this mold.[33] While there are some (published) arguments in favor of d *as opposed to* r, there seem to be no arguments in the literature which favor d over the alternatives l, $\mathfrak{r}$, or s.[34] Moreover, as I will show in the next section, only one of the two popular arguments in favor of d *as opposed to* r is *at all* compelling. In contrast, several *general* arguments in favor of r, l, and $\mathfrak{r}$ *have* appeared in the literature.[35] It is precisely this kind of *general* argument that is needed to undergird the use of one particular relevance measure rather than any other.

In the next section, I will examine two recent arguments in favor of the difference measure d *as opposed to* the log-ratio measure r. While one of these arguments holds some promise of adjudicating between d and r (in favor of d), I will argue that neither of them will help to adjudicate between d and l, or between d and $\mathfrak{r}$, or d and s. As a result, defenders of the difference measure will need to do further logical work to complete their enthymematic confirmation-theoretic arguments.

---

of Horwich's (1982) account of the confirmational significance of evidential diversity. In §3.3.2 below, I present a more charitable alternative reconstruction of Horwich's account of CSED [one *not* reliant on (5)], which is *not* sensitive to choice of measure.

[33]Kaplan (1996) offers several criticisms of Bayesian confirmation theory which presuppose the adequacy of the difference measure d. He then suggests (76, footnote 73) that all of his criticisms will also go through for all other relevance measures that have been proposed in the literature. But, one of his criticisms (84, footnote 86) does *not* apply to measure r.

[34]See Eells and Fitelson (2000, 2001) for some good reasons to prefer d (and l) over s (and $\mathfrak{r}$).

[35]Milne (1996) argues that r is "the one true measure of confirmation." Good (1984), Heckerman (1988), and Schum (1994) all give general arguments in favor of l. And, Carnap (1962, §67) gives a general argument in favor of $\mathfrak{r}$. In §3.1 below, I briefly discuss the arguments of Milne (1996), Good (1984), and Carnap (1962, §67), and in §3.2, I provide my own argument for l.

2.3 Two Arguments *Against* r

2.3.1 "Deductive Insensitivity" Argument Against r

Rosenkrantz (1981) and Gillies (1986) point out the following fact about r:

(6) If $H \vDash E$, then $r(H, E \mid K) = r(H \,\&\, X, E \mid K)$, for any X.

Informally, (6) says that, *in the case of deductive evidence,* $r(H, E \mid K)$ *does not depend on the logical strength of* H. Gillies (1986) uses (6) as an argument against r, and in favor of the difference measure d. Rosenkrantz (1981) uses (6) as an argument against r, but he cautiously notes that *neither* d *nor* l satisfies (6). It is easy to show that neither $\mathfrak{r}$ nor s has property (6) either (proofs omitted).

I think Gillies (1986, page 112) pinpoints rather well what is so peculiar and undesirable about (6) when he explains that:

> On the Bayesian, or, indeed, on any inductivist position, the more a hypothesis H goes beyond [deductive] evidence E, the less H is supported by E. We have seen [in (6)] that r lacks this property that is essential for a Bayesian measure of support.

I agree with Gillies and Rosenkrantz that this argument provides a somewhat compelling reason to abandon r in favor of *either* d, l, s, *or* $\mathfrak{r}$.[36] But, it says

[36]Milne (1996) argues that, in the case of deductive evidence, r's (alleged) mishandling of the problem of irrelevant conjunction [which stems from r's violation of (6)] is *not* a reason to reject r. Indeed, Milne (1996) characterizes the traditional, deductive problem of irrelevant conjunction as a "wretched shibboleth." I am somewhat sympathetic to Milne here. Like Milne, I think one should probably not place *too* much emphasis on deductive cases (after all, what we're after here is a measure of *in*ductive support). And, like Milne, I also think that the standard Bayesian accounts of the traditional, deductive problem of irrelevant conjunction are wrongheaded. See §2.2.2.1 (and footnote 28) for my analysis of the problem of irrelevant conjunction. There, I show that the *in*ductive version of this problem (when properly analyzed) *does* expose a more serious weakness of the measure r. *Pace* Milne (1996), I think there are *lots* of compelling reasons to reject r. See §3.1.2, §3.2.2, and Eells and Fitelson (2001) for just of few of these. And, see §3.1.1 for my analysis (and critique) of Milne's (1996) *desideratum/explicatum* argument in favor of r.

nothing about *which* of d, l, s, or $\mathfrak{r}$ should be adopted. So, this argument does not suffice to shore-up all of the measure sensitive arguments we have seen. Hence, it does not constitute a complete resolution of the problem of measure sensitivity.

2.3.2 "Unintuitive Confirmation" Argument Against r

Several recent authors, including Sober (1994b) and Schum (1994), have criticized r on the grounds that r sanctions "unintuitive" quantitative judgments about degree of confirmation in various (hypothetical) numerical examples.[37] For instance, Sober (1994b) asks us to consider a hypothetical case involving a single collection of evidence E, and two hypotheses H_1 and H_2 (where, K is taken to be *tautologous*, and is thus suppressed) such that:

$$\begin{array}{ll} \Pr(H_1 \mid E) = 0.9 & \Pr(H_1) = 0.09 \\ \Pr(H_2 \mid E) = 0.0009 & \Pr(H_2) = 0.00009 \end{array}$$

In such a case, we have the following pair of probabilistic facts:

$$(\dagger) \qquad \begin{array}{l} d(H_1, E) = 0.81 \gg d(H_2, E) = 0.00081 \\ r(H_1, E) = \log(10) = r(H_2, E) \end{array}$$

It is then argued, by proponents of d, that (†) exposes a highly "unintuitive" feature of r, since this is case in which — "intuitively" — E confirms H_1 to a greater degree than E confirms H_2. But, according to r, E confirms both H_1 and H_2 to exactly the same degree. Therefore, this example is purported to *rule out* r (but *not* d, since d gets the "intuitively correct" answer here).

I am not too worried about (†), for two reasons. First, (†) can be only a reason to favor the difference measure over the ratio measure (or vice versa[38]); it has

[37] Sober (1994b) borrows this line of criticism from Ellery Eells. Eells (personal communication) has voiced numerical examples of various kinds to illustrate the "unintuitive" consequences of r.

[38] It has been argued by Schlesinger (1995) [and Pollard (1999)] that parallel arguments can

little or no bearing on the relative adequacy of either l, s, or $\mathfrak{r}$. It is clear from the definitions of the measures that Carnap's $\mathfrak{r}$ and Christensen's s are bound to *agree* with d's "intuitive" answer in such cases. Hence, $\mathfrak{r}$ and s are *immune* from the "unintuitive confirmation" criticism. Moreover, the log-likelihood ratio measure l certainly *could* agree with the "intuitively" correct judgments in such cases (depending on how the details needed to fix the *likelihoods* get filled-in). Indeed, Schum (1994, Ch. 5) argues nicely that the log-likelihood ratio measure l is largely immune to the kinds of "scaling effects" exhibited by r and d in (†). Unfortunately, neither Eells nor Sober (1994b) nor Schlesinger (1995) considers how the measures l, s, and $\mathfrak{r}$ might cope with their alleged counter-examples.

Second, there seems to be little or no *independent* support offered for the crucial premise of this argument. The argument is persuasive only if it is granted that the *intuitive* degree to which E confirms H_1 *is* greater than the *intuitive* degree to which E confirms H_2. The only reason that I have seen offered in support of this claim (*e.g.,* Sober (1994b)) is that $d(H_1, E) \gg d(H_2, E)$. But, this just seems to *beg the question*; it simply presupposes that the *intuitive* amount to which E confirms H *is* accurately gauged by the *difference* measure, and *not* by the *ratio* measure (or, by some other measure altogether). What we need here are *independent reasons* for believing *precisely this*!

2.4 Summary of Results So Far

We have discussed three measure sensitive arguments which are aimed at showing that certain relevance measures are *in*adequate, and we have seen five measure sensitive arguments which presuppose the *superiority* of certain relevance measures

be run "backward" *against* d and *in favor of* r. Schlesinger (1995) describes a class of examples in which the difference measure seems to give the "unintuitive" answer (and, where the key probabilistic facts are analogous to (†)). Schlesinger's examples drive home the point that the *philosophical* conclusions one draws from hypothetical, *numerical* examples of these kinds will depend crucially on what one takes the *"intuitive"* answers to be in the first place. See below.

over others. Table 2 summarizes the arguments which presuppose that certain relevance measures are superior to others, and Table 3 summarizes the arguments against various relevance measures. These tables serve as a handy reference on the measure sensitivity problem in Bayesian confirmation theory.

	Is $\mathcal{A}$ valid wrt the measure:				
Name and Section of Argument $\mathcal{A}$	d?	r?	l?	$\mathfrak{r}$?	s?
Rosenkrantz on "Irrelevant Conjunction" (See §2.2.2 and Appendix §A.2 for discussion)	Yes	No	No	Yes	Yes
Earman on "Irrelevant Conjunction" (See §2.2.2 for discussion)	Yes	No	Yes	Yes	Yes
Eells on the Grue Paradox (See §2.2.3 and Appendix §A.4 for discussion)	Yes	No	No	Yes	Yes
Sober on the Grue Paradox (See §2.2.3 for discussion)	Yes	No	Yes	Yes	Yes
Horwich *et al.* on Ravens *&* Variety (See §2.2.4 and Appendix §A.5 for discussion)	Yes	Yes	Yes	No	No

Table 2: Five arguments which presuppose the *superiority* of certain measures.

	Is $\mathcal{A}$ valid wrt the measure:				
Name and Section of Argument $\mathcal{A}$	d?	r?	l?	$\mathfrak{r}$?	s?
Gillies's Popper-Miller Argument (See §2.2.1 and Appendix §A.1 for discussion)	Yes	No	No	Yes	Yes
"Deductive Insensitivity" Argument (See §2.3.1 for discussion)	No	Yes	No	No	No
"Unintuitive Confirmation" Argument (See §2.3.2 for discussion)	No	Yes[39]	No	No	No

Table 3: Three arguments designed to show the *inadequacy* of certain measures.

[39] As I explain in § 2.3.2, I do *not* think this argument is compelling, even when aimed against r. But, to be charitable, I will grant that it is, at least, *valid* when aimed against r.

2.5 Where Do We Go From Here?

In this chapter, I have shown that many well-known arguments in quantitative Bayesian confirmation theory are valid only if the difference measure d is to be preferred over other relevance measures (at least, in the confirmational contexts in question). I have also shown that there are some good reasons to prefer d over the log-ratio measure r. Unfortunately, like Rosenkrantz (1981), I have found *no* compelling reasons offered in the literature to prefer d over the log-*likelihood* ratio measure l (or Carnap's relevance measure $\mathfrak{r}$, or Christensen's measure s). As a result, philosophers like Gillies, Rosenkrantz, and Eells, whose arguments presuppose that d is preferable to l, s, and $\mathfrak{r}$ seem compelled to produce some *justification* for using d, rather than l, s, or $\mathfrak{r}$, to measure degree of confirmation.

In general, there seem to be two viable strategies for coping with the problem of measure sensitivity. The first strategy is to simply *avoid* the problem entirely, by making sure that one's quantitative confirmation-theoretic arguments are *robust* (*i.e., in*sensitive to choice of measure of confirmation).[40] On the other hand, if plausible robust arguments can *not* be found in some context, then one should feel compelled to give reasons why one's chosen relevance measure (or class of relevance measures) $\mathfrak{c}^*$ should be preferred over other relevance measures, the use of which would render one's argument invalid.

Ideally, it would be nice to see general, *desideratum/explicatum* arguments which rule out all but a relatively small class of ordinally equivalent measures of confirmation [*i.e.*, arguments like those given by Carnap (1962, §67), Good (1984), Heckerman (1988), and Milne (1996)]. Such arguments would also have the virtue of contributing in a substantive way to the *theoretical underpinning* of

[40]This *can* be done in some contexts. For instance, in Fitelson (2001a) (and in the next chapter), I outline a new, robust Bayesian resolution of the problem of evidential diversity. And, Maher (1999) gives a new, robust Bayesian resolution of the Ravens Paradox, based on Carnapian inductive logic. I doubt, however, that plausible, robust Bayesian accounts can *always* be found.

quantitative Bayesian confirmation theory. In the next chapter, I briefly discuss a few such arguments that have appeared in the literature, and I describe several (novel) independent ways of narrowing the field of measures. Ultimately, this will lead to an almost unique solution of the problem of measure sensitivity.

Chapter 3

Independent Evidence, Measures of Confirmation, and The Value of Evidential Diversity[41]

He who has heard the same thing told by 12,000 eye-witnesses
has only 12,000 probabilities, which are equal to one strong
probability, which is far from certain.
— Voltaire

In this chapter, I will (*i*) survey and critique a few existing attempts to resolve the problem of measure sensitivity, and (*ii*) describe several simple and novel ways (of my own) of narrowing the field of relevance measures. Along the way, I will (*iii*) outline a new Bayesian account of independent evidence, which will be applied to both the problem of measure sensitivity and the problem of evidential diversity. Finally, I will (*iv*) compare my account of evidential diversity with several other Bayesian approaches that have appeared in the literature.

3.1 Three Existing Attempts to Solve the Problem of Measure Sensitivity

In this section, I will briefly discuss arguments of Milne (1996), Carnap (1962), and Good (1984), which, if cogent, would resolve the problem of measure sensitivity once and for all, by establishing one relevance measure as "the one true Bayesian

[41] Much of the material in this chapter appears in Fitelson (1996) and Fitelson (2001a).

measure of confirmation." Unfortunately, none of these existing arguments proves to be persuasive. In subsequent sections, I will describe some (novel) simple and intuitive ways of drastically narrowing the field of competing measures.

3.1.1 Milne's Reductionistic Argument for r

Milne (1996) shows that the only measure (up to ordinal equivalence) which satisfies all five of the following desiderata is the log-ratio measure r.[42]

(7) $\mathfrak{c}(H, E \mid K) > 0$ when $\Pr(H \mid E \,\&\, K) > \Pr(H \mid K)$; $\mathfrak{c}(H, E \mid K) < 0$ when $\Pr(H \mid E \,\&\, K) < \Pr(H \mid K)$; $\mathfrak{c}(H, E \mid K) = 0$ when $\Pr(H \mid E \,\&\, K) = \Pr(H \mid K)$. [In other words, $\mathfrak{c}(H, E \mid K)$ must satisfy $\mathcal{R}$.]

(8) $\mathfrak{c}(H, E \mid K)$ is some function of the values $\Pr(\cdot \mid K)$ and $\Pr(\cdot \mid \cdot \& K)$ assumed on the at most sixteen truth-functional combinations of E and H.

(9) If $\Pr(E \mid H \,\&\, K) < \Pr(E' \mid H \,\&\, K)$ and $\Pr(E \mid K) = \Pr(E' \mid K)$ then $\mathfrak{c}(H, E \mid K) \leq \mathfrak{c}(H, E' \mid K)$; if $\Pr(E \mid H \,\&\, K) = \Pr(E' \mid H \,\&\, K)$ and $\Pr(E \mid K) < \Pr(E' \mid K)$ then $\mathfrak{c}(H, E \mid K) \geq \mathfrak{c}(H, E' \mid K)$.

(10) $\mathfrak{c}(H, E_1 \,\&\, E_2 \mid K) - \mathfrak{c}(H, E_1 \,\&\, E_3 \mid K)$ is fully determined by $\mathfrak{c}(H, E_1 \mid K)$ and $\mathfrak{c}(H, E_2 \mid K \,\&\, E_1) - \mathfrak{c}(H, E_3 \mid K \,\&\, E_1)$; if $\mathfrak{c}(H, E_1 \,\&\, E_2 \mid K) = 0$ then $\mathfrak{c}(H, E_1 \mid K) + \mathfrak{c}(H, E_2 \mid K \,\&\, E_1) = 0$.

(11) If $\Pr(E \mid H \,\&\, K) = \Pr(E \mid H' \,\&\, K)$ then $\mathfrak{c}(H, E \mid K) = \mathfrak{c}(H', E \mid K)$.

Milne's argument has several flaws. First, in addition to (7)–(11), Milne's argument *implicitly* requires that the probability function Pr (and, hence, the "*explicatum*" $\mathfrak{c}$) satisfy some rather strong, unmotivated, and unintuitive constraints.[43]

[42] I have taken the liberty of translating Milne's desiderata into our notation.

[43] Like similar arguments of Cox (1961), Good (1984), and Heckerman (1988), Milne's argument makes use of certain theorems in the theory of functional equations, which force the probability function Pr (and, hence, the spaces over which the measure $\mathfrak{c}$ is defined) to satisfy various kinds of *continuity* conditions (and other constraints which force the underlying function spaces in question to be infinite in various ways). These assumptions are discussed in detail (and shown to be implausible, at least on an epistemic reading of Pr) by Halpern (1999a, 1999b). I won't bother

Second, and more importantly, not all of Milne's "desiderata" (7)–(11) are philosophically well motivated. Milne's (7)–(9) are rather uncontroversial; they are satisfied by almost every relevance measure that's been proposed in the literature (including our five: d, r, l, s, and $\mathfrak{r}$; see §A.10 for proofs). However, (10) and (11) are far more controversial (and far less intuitive) than Milne would have us believe. For one thing, Milne (1996, p. 22) seems unaware that there are several proposed relevance measures which violate his desideratum (10), when he says:

> We may note that, like (9), desideratum (10) is a consequence of the most commonly used measure of confirmation $d(H, E \mid K) = \Pr(H \mid E \& K) - \Pr(H \mid K)$. Any substantive reason for rejecting either (9) or (10) cuts a swath through the literature on probabilistic confirmation theory.

This quote suggests that Milne views the competition among relevance measures to be only between d and r.[44] As we know, many other relevance measures have been proposed and defended. Indeed, Christensen (1999) has argued in favor of s, as opposed to d, r, and l. And, Carnap (1962) explicitly provides reasons to favor $\mathfrak{r}$ over r. Unfortunately, both s and $\mathfrak{r}$ *violate* (10) (see §A.10), and Milne provides no argument for (10). In the absence of such an argument, Milne has given us no reason to favor r over s or $\mathfrak{r}$ [or any other measure which violates (10)].

Finally, there is Milne's desideratum (11). Strangely, this desideratum *alone* is enough to single r out of the five competing relevance measures we have been

to discuss these *technical* shortcomings in Milne's argument in detail, since this would require a rather extensive mathematical digression (which Halpern handles beautifully), and because the *philosophical* problems with Milne's argument are, I think, more interesting. However, I will say that it is somewhat misleading for Milne to claim that the only measure which satisfies (7)–(11) is r, when, in fact, Milne's argument requires many more mathematical "*desiderata*" than just these five. Strictly speaking, all Milne has *really* shown is that the only measure (up to ordinal equivalence) that satisfies (7)–(11) — *assuming that the underlying probability spaces and function spaces satisfy lots of other strong (and implausible) mathematical constraints* — is r. This seems a far cry from showing that r is "the one true measure of confirmation."

[44] To be fair to Milne, the log-likelihood ratio measure l *also* satisfies (10) (proof omitted). He does not mention this, and, even on this more charitable reading, this quote seems to indicate that Milne is only taking seriously d, r, and l as candidates for "the one true measure of confirmation." He seems unaware that measures like s and $\mathfrak{r}$, which violate (10), need to be taken seriously.

talking about. That is, of the five measures d, r, l, s, and $\mathfrak{r}$, *only r satisfies* (11).[45] So, it is clearly (11) that is the most controversial of Milne's five desiderata. The good news is that Milne is well aware of this, and he spends some time trying to philosophically motivate (11). The bad news is that Milne's argument for (11) rests on a conflation of relational and non-relational notions of evidential support, and is consequently unsound. Milne (1996, p. 22) seems to think that (11) is a consequence of the following, which he calls the Likelihood Principle (LP):

> In comparing the evidential bearing (relative to background knowledge K) of E on the hypotheses H and H' we need consider only $\Pr(E \mid H \,\&\, K)$ and $\Pr(E \mid H' \,\&\, K)$.

Unfortunately, there is an implicit (and spurious) assumption in Milne's argument from (LP) to (11). The likelihood principle is intended to tell us when evidence E favors one hypothesis H over another hypothesis H'. According to the likelihood principle, E favors H over H' relative to K iff $\Pr(E \mid H \,\&\, K) > \Pr(E \mid H' \,\&\, K)$. Notice that "$E$ favors H over H' relative to K" is a *four*-place relation. It is far from obvious that the four-place favoring relation can (or should) be reduced to the three-place confirmation relation. That is, the following reductionistic presupposition in Milne's reasoning from (LP) to (11) is far from obvious:

(11′) E favors H over H' relative to K iff $\mathfrak{c}(H, E \mid K) > \mathfrak{c}(H', E \mid K)$.

Likelihoodists [*e.g.*, Royall (1997)] would certainly *reject* (11′), and with it the move from (LP) to (11). After all, one of the main reasons for making the move from "confirmation" to "favoring" is that doing so allows us to avoid having to

[45]Indeed, as far as I know, r is the only measure that has been proposed or defended in the literature on confirmation theory which satisfies (11). Given this fact, one wonders why Milne bothers with (10) in the first place. I suppose he thought he "needed" (10), in the context of the particular mathematical proof strategy he happened to choose. This is yet another reason to abandon his rather complex and subtle mathematical argument in favor of a much simpler argument which is sufficient to rule-out all measures that have actually been proposed and defended.

worry about prior probabilities (*i.e.*, $\Pr(H \mid K)$) and/or likelihoods of "catch all" logical negations (*i.e.*, $\Pr(E \mid \bar{H} \;\&\; K)$, where $\bar{H}$ is the logical negation of some concrete hypothesis with a well understood and precise likelihood). Likelihoodists are notorious for being quite skeptical about the objectivity or meaningfulness of claims about confirmation. Many of the proponents of (LP) think that, in the vast majority of cases, $\mathfrak{c}(H, E \mid K)$ is either utterly subjective or ill-defined (or both), since it will (generally) depend on prior probabilities and/or likelihoods of "catch all" logical negations.[46] So, it seems quite odd that Milne would try to use (LP) in an argument for a measure of (non-relational) confirmation.

The problem here can be understood as a difficulty that arises from trying to reduce an inherently relational notion to a non-relational notion. Consider the following analogy with physics. One might claim (as Newton did) that relative velocities can be reduced to (or defined in terms of) absolute velocities with respect to the æther. Of course, someone who does not believe in the existence of the æther will simply reject such a reductive definition of relational velocity. Analogously, Milne seems to be claiming that the relational evidential notion of favoring can be reduced to (or defined in terms of) the non-relational notion of confirmation (with respect, if you like, to the "logical æther"). Likelihoodists, of course, *reject* the existence (at least, the *objective* existence) of the kind of "logical æther" (*viz.*, *a priori* probabilities) needed for such a reduction to go through.

There is one special case in which Milne's reductionistic principle (and his move from (LP) to (11) to r) makes sense. That is the case in which $H' = \bar{H}$ (*i.e.*, when H' is *identical* to the logical negation of H). In this case, it is true that E favors H over H' relative to K iff $r(H, E \mid K) > r(H', E \mid K)$. But, this cannot be used as a reason to favor r over any other relevance measure, since in this case, we will have

[46] Most notably, Royall (1997, §1.5) provides a general argument *against* thinking of evidential support in a non-relational way. In particular, Royall (1997, pp. 9–11) gives an argument *against* Milne's measure $r(H, E \mid K)$! This is ironic, since Milne would have us believe that advocates of (LP), like Royall, are (somehow) committed to *accepting* r as a measure of support.

E favors H over H' relative to K iff $\mathfrak{c}(H, E \mid K) > 0$, for *any* relevance measure $\mathfrak{c}$. That is, in this special case, favoring and confirmation *do* amount to the same thing. The problem arises when H' is *not* identical to the logical negation of H (*i.e.*, when H' is a concrete alternative hypothesis with a well-defined likelihood on E). It is precisely *those* cases in which the Likelihoodist and the Bayesian will disagree about how the problem should be analyzed.[47] Table 4 summarizes the main results from this section (see §A.10 for selected proofs):

	Is $\mathcal{D}$ satisfied by the measure:				
Milne's Desideratum $\mathcal{D}$	d?	r?	l?	$\mathfrak{r}$?	s?
(7)	YES	YES	YES	YES	YES
(8)	YES	YES	YES	YES	YES
(9)	YES	YES	YES	YES	YES
(10)	YES	YES	YES	NO	NO
(11)	NO	YES	NO	NO	NO

Table 4: Summary of results concerning Milne's argument for r.

[47] Ironically, Milne comes very close to adopting the likelihood-ratio measure here. Essentially, Milne is recommending the likelihood ratio as an adequate measure of degree of *favoring*, but *not* as an adequate measure of degree of *confirmation*. I will argue below that likelihood ratios should be used *both* to measure degree of favoring [as suggested in Royall (1997)] *and* to measure degree of confirmation. I.J. Good has been making the same suggestion (for various reasons) for many years. Moreover, Good (1983, pp. 36–37) is very careful to distinguish the cases in which H and H' form a logical partition, and the cases in which they do not. This is the key to seeing why (and how) the likelihood ratio gives the right answer in *both* kinds of cases.

3.1.2 Carnap's Symmetry Argument for $\mathfrak{r}$

Carnap (1962, §67) shows that $\mathfrak{r}$ exhibits all four of the following symmetries.

$$\mathfrak{c}(H, E \mid K) = \mathfrak{c}(E, H \mid K) \tag{12}$$

$$\mathfrak{c}(H, E \mid K) = -\mathfrak{c}(H, \bar{E} \mid K) \tag{13}$$

$$\mathfrak{c}(H, E \mid K) = -\mathfrak{c}(\bar{H}, E \mid K) \tag{14}$$

$$\mathfrak{c}(H, E \mid K) = \mathfrak{c}(\bar{H}, \bar{E} \mid K) \tag{15}$$

Carnap seems aware that neither d nor r satisfies all four of these constraints. And, as Eells and Fitelson (2001) show, it turns out that neither s nor l satisfies all of these symmetries either. So, Carnap has provided, essentially, a list of symmetry desiderata that rule-out all of our five relevance measures, except his $\mathfrak{r}$. If Carnap had provided good reasons for thinking that $\mathfrak{c}$ *should* satisfy all of (12)–(15), then he would have a had a pretty compelling argument in favor of $\mathfrak{r}$.[48]

Table 5 — reproduced from Eells and Fitelson (2001) — shows which measures satisfy which of Carnap's symmetry properties (12)–(15) [all proofs have been omitted here, but they can be found in Eells and Fitelson (2001)].

	Is $\mathfrak{D}$ satisfied by the measure:				
Carnap's Desideratum $\mathfrak{D}$	d?	r?	l?	$\mathfrak{r}$?	s?
(12)	No	Yes	No	Yes	No
(13)	No	No	No	Yes	Yes
(14)	Yes	No	Yes	Yes	Yes
(15)	No	No	No	Yes	Yes

Table 5: Summary of results concerning Carnap's argument for $\mathfrak{r}$.

[48] It's not entirely clear *why* Carnap thinks it is a *good thing* for a relevance measure $\mathfrak{c}$ to satisfy (12)–(15). At times, it seems he's thinking about pragmatic factors like mathematical elegance, beauty, or ease of computation. But, as Eells and Fitelson (2001) argue, this does not seem to jibe with firm intuitions about how measures of degree of evidential support ought to behave.

As we have explained in Eells and Fitelson (2001), we think that symmetry considerations can be relevant and useful in this context. However, we think that too much symmetry (as exhibited especially by $\mathfrak{r}$) is a *bad* thing for a measure of degree of confirmation or support. Consider the following example [again, taken from Eells and Fitelson (2001)]:

> A card is randomly drawn from a standard deck. Let E be the evidence that the card is the seven of spades, and let H be the hypothesis that the card is black. We take it to be intuitively clear that E is not only conclusive, but also strong, evidence in favor of H, whereas: $\bar{E}$ (that the card drawn is *not* the seven of spades) is close to useless, or close to "informationless," with regard to the color of the card. ... With initial uncertainty about the value of the card, we consider the seven of spades, as evidence, to be more highly informative and confirmatory of the blackness of the card, as hypothesis, than the blackness of the card, as evidence, is for the card's being the seven of spades in particular.

In other words, this simple example shows clearly that the symmetry conditions (12) and (13) are *not* generally satisfied by an adequate measure $\mathfrak{c}$ of degree of support.[49] This gives us a simple and intuitive way of *ruling-out* Carnap's $\mathfrak{r}$ as well as Milne's r and Christensen's s. That is, two simple considerations of symmetry allow us to narrow the field to d and l. Later in this chapter, another, independent way of narrowing the field to d and l will be reported. Then, a final adjudication between d and l (in favor of l) will be presented. But, first, we'll take a quick look at Good's "best explicatum" argument for the log-likelihood-ratio measure l.

[49]The conclusiveness feature of the examples (that E logically implies H) is not what is at the heart of the counterexample. To see this, simply consider a modification of the examples where E is a report of suit/rank, respectively, of very reliable, but fallible, assistant.

3.1.3 Good's "Best Explicatum" Argument for l

Good (1984) claims to show that l is the best explicatum for "weight of evidence." He seems, specifically, to be interested in showing that l is a better measure of degree of confirmation than either d or r, when he reports that (in our notation):

> ... One reason for writing the present note is that the demonstration in Good (1968) has been overlooked by several philosophers of science. For example, Rosenkrantz (1981) says "I know of no compelling considerations that adjudicate between the difference measure d and Good's weight of evidence l." Also, Horwich (1982) mentions only r and d as potential explicata ...

Good goes on to provide the following two 'compelling' desiderata for an adequate measure $\mathfrak{c}$ of degree of confirmation.

(16) $\mathfrak{c}(H, E \mid K)$ must be a function f only of $\Pr(E \mid H \& K)$ and $\Pr(E \mid \bar{H} \& K)$.

(17) This function $f[\Pr(E \mid H \& K), \Pr(E \mid \bar{H} \& K)]$, together with $\Pr(H \mid K)$, must mathematically determine $\Pr(H \mid E \& K)$.

Like Milne's argument for r, Good's argument (if cogent) would establish l as "the one true measure of confirmation" (up to ordinal equivalence). And, like Milne's argument, Good's argument has two main problems. First, it makes use of the same kinds of functional equational analyses that require far stronger mathematical assumptions about Pr [and, hence, the explicatum $\mathfrak{c}$ satisfying (16) and (17)] than Good would have us believe [see footnote 43 and Halpern (1999a, 1999b)]. Moreover, Good provides far too little argumentation in support of his desiderata (16) and (17). For instance, consider what Good has to say in support of (16):

> ... Note that the first desideratum (16) implies that the weight of evidence in favor of H provided by E does not depend on the prior

> probability of H. This prior probability might be large or small depending on previous or other evidence, not of course on E.

This is highly misleading, to say the least. To see why, note that l can be rewritten in the following (*numerically* equivalent) way, which (in a naïve, syntactical sense) seems to suggest that l "depends only on $\Pr(H \mid E \,\&\, K)$ and $\Pr(H \mid K)$."

$$l(H, E \mid K) = \log\left[\frac{\Pr(H \mid E \,\&\, K) \cdot [1 - \Pr(H \mid K)]}{[1 - \Pr(H \mid E \,\&\, K)] \cdot \Pr(H \mid K)}\right]$$

On this way of writing l, it appears that l *does* "depend on the prior probability of H [*i.e.*, $\Pr(H \mid K)$]." Moreover, the measure r can be rewritten as follows, which seems to suggest that r does *not* "depend on the prior probability of H."

$$r(H, E \mid K) = \log\left[\frac{\Pr(E \mid H \,\&\, K)}{\Pr(E \mid K)}\right]$$

What, then, are we to make of Good's talk of "dependence" on the prior probability of H? I'm not really sure. It seems that Good is simply imposing question-begging, *syntactical* requirements on the functional form of $\mathfrak{c}$. After all, *no* relevance measure can be semantically (or algebraically) *generally independent* of the prior probability of H. This kind of general invariance under perturbations of $\Pr(H \mid K)$ would force $\mathfrak{c}(H, E \mid K)$ to *violate* $\mathcal{R}$. So, I'm afraid that Good's argument for l is even weaker than Milne's argument for r. In the next section, I will develop a Bayesian account of confirmationally independent evidence regarding a hypothesis. As we'll see, this *semantical* notion of independence *can* be used to provide a compelling argument for l as "the one true measure of confirmation."

3.2 A Bayesian Account of Independent Evidence

If $\mathfrak{c}(H, E \mid K)$ is an adequate measure of the degree to which E confirms H relative to K, then $\mathfrak{c}(H, E \mid K)$ will in general vary depending on what the background evidence K is. For example, let H be the hypothesis that something is wrong with a computer and let E be the evidence that nothing happens when the computer is turned on. If the background evidence K includes facts such as that the computer is plugged in, then E will confirm H relative to K; on the other hand, if K specifies that the computer is not plugged in and that it needs to be plugged in to work, then E will *not* confirm H relative to K.

When we want to consider how degree of confirmation varies with changing background evidence, we will use the conditional notation $\mathfrak{c}(H, E_1 \mid E_2)$ to denote the degree to which E_1 confirms H (according to $\mathfrak{c}$), given that E_2 is part of our background evidence.[50] And, we will use the unconditional notation $\mathfrak{c}(H, E_1)$ to denote the degree to which E_1 confirms H (according to $\mathfrak{c}$), *not* conditional on E_2 being part of our background evidence. The point of the preceding paragraph is that, for any adequate measure of confirmation $\mathfrak{c}$, there are cases in which $\mathfrak{c}(H, E_1 \mid E_2) \neq \mathfrak{c}(H, E_1)$. When this happens, we say that E_1 is *confirmationally dependent on* E_2 *regarding* H *according to* $\mathfrak{c}$. Conversely, if $\mathfrak{c}(H, E_1 \mid E_2) = \mathfrak{c}(H, E_1)$ then we say that E_1 is *confirmationally independent of* E_2 *regarding* H *according to* $\mathfrak{c}$. If both $\mathfrak{c}(H, E_1 \mid E_2) = \mathfrak{c}(H, E_1)$, and $\mathfrak{c}(H, E_2 \mid E_1) = \mathfrak{c}(H, E_2)$, then we say that E_1 and E_2 are *mutually confirmationally independent (or, simply, independent) regarding* H *according to* $\mathfrak{c}$. As it turns out, C.S. Peirce (1878) had some interesting things to say about confirmational independence. In the next section, we will use Peirce's early intuitions about independent evidence to lay the groundwork for our

[50] There may be other background evidence besides E_2 in a confirmational context. However, this additional background evidence will be *held fixed* in the confirmational comparisons we do to determine whether E_1 and E_2 are dependent or independent regarding H in that context. So, there is no need to indicate this additional background evidence explicitly. As such, I will, for simplicity, hereafter suppress the (full) background evidence K from my notation.

own, Bayesian account.

3.2.1 The Fundamental Peirceian Desiderata

In his essay "The Probability of Induction," C.S. Peirce articulates several fundamental intuitions concerning the nature of independent inductive support. Consider the following important excerpt from Peirce (1878, my brackets):

> ... two arguments which are entirely independent, neither weakening nor strengthening the other, ought, when they concur, to produce a[n intensity of] belief equal to the sum of the intensities of belief which either would produce separately.

Two crucial intuitions about independent inductive support are contained in this quote. First, there is the intuition that two pieces of evidence E_1 and E_2 provide *independent* inductive support for a hypothesis H *if the degree to which* E_1 *supports* H *does not depend on whether* E_2 *is part of our background evidence (and vice versa).* In our confirmation-theoretic framework, we will take this intuition (already discussed briefly in the previous section) onboard as our official *definition* of (mutual) confirmational independence regarding a hypothesis:

> **Definition**. E_1 and E_2 are *confirmationally independent regarding* H *according to* $\mathfrak{c}$ iff $\mathfrak{c}(H, E_1 \mid E_2) = \mathfrak{c}(H, E_1)$ and $\mathfrak{c}(H, E_2 \mid E_1) = \mathfrak{c}(H, E_2)$.[51]

[51] James Joyce and Patrick Maher (private communications) have both voiced concerns about whether this is an accurate reading of Peirce. They worry that Peirce is talking in this passage *not* about the degree of incremental confirmation $\mathfrak{c}(H, E)$, but about the posterior probability $\Pr(H \mid E)$. While this may be true [as a psychological fact about Peirce — although Good (1983) and Schum (1994) seem to think otherwise], this would not undermine the cogency of my subsequent arguments. For, I intend only to take Peirce's somewhat vague statements as a historical inspiration for my own account. However, it is interesting to note that, if Peirce *is* talking about the posterior probability here, then his requirement of additivity in cases where $\Pr(H \mid E_1 \& E_2) = \Pr(H \mid E_1) = \Pr(H \mid E_2)$ makes no sense, since in such cases: $\Pr(H \mid E_1 \& E_2) = \Pr(H \mid E_1) = \Pr(H \mid E_2) \neq \Pr(H \mid E_1) + \Pr(H \mid E_2)$. So, I consider my reading of this passage to be a rather charitable one. Moreover, the definition of confirmational independence I am adopting is a natural and (pre-theoretically) intuitive one. Interestingly, many researchers in

The second intuition expressed by Peirce in this passage is that the joint support provided by two pieces of independent evidence should be *additive*. In our confirmation theoretic framework, this gets unpacked as follows:

($\mathcal{A}'$) If E_1 and E_2 are confirmationally independent regarding H according to $\mathfrak{c}$, then $\mathfrak{c}(H, E_1 \;\&\; E_2) = \mathfrak{c}(H, E_1) + \mathfrak{c}(H, E_2)$.

Strictly speaking, we should weaken $\mathcal{A}'$ to require only that $\mathfrak{c}(H, E_1 \;\&\; E_2)$ be *some* (symmetric) isotone function f of $\mathfrak{c}(H, E_1)$ and $\mathfrak{c}(H, E_2)$, where f is additive in *some* (isotonically) transformed space.[52] The point is that, if E_1 and E_2 are confirmationally independent regarding H according to $\mathfrak{c}$, then $\mathfrak{c}(H, E_1 \& E_2)$ should depend *only* (and, in some isotonically transformed space, *linearly*) on $\mathfrak{c}(H, E_1)$ and $\mathfrak{c}(H, E_2)$, without any extra "interaction terms." This leads to the following refinement of the second basic Peirceian intuition:

($\mathcal{A}$) There exists some (symmetric) isotone function f such that, for all E_1, E_2, and H, if E_1 and E_2 are confirmationally independent regarding H according to $\mathfrak{c}$, then $\mathfrak{c}(H, E_1 \& E_2) = f[\mathfrak{c}(H, E_1), \mathfrak{c}(H, E_2)]$, where f is additive in some (isotonically) transformed space.

The following theorem states that each of our five Bayesian relevance measures — *except* $\mathfrak{r}$ *and* s — satisfies $\mathcal{A}$ (see the Appendix for proofs of all theorems).

Theorem 6. *Each of the measures d, r, and l satisfies $\mathcal{A}$, but s and $\mathfrak{r}$ violate $\mathcal{A}$.*

So, at this most basic level, the three most popular varieties of quantitative Bayesian confirmation theory are in agreement about the nature of independent

artificial intelligence have adopted the very same definition. They call it 'modularity' — see, *e.g.*, Horvitz and Heckerman (1986) and Heckerman (1988).

[52] As Peirce did, I prefer to have f be $+$. So, I have defined r and l using *logarithms* (see footnote 14). If we were to drop the logarithms in our definitions of r and l, then we would have $f = \cdot$ for the ratio measures r and l, but $f = +$ for the difference measure d. See Heckerman (1988) for more on the kind of linear decomposability that is at the heart of desideratum $\mathcal{A}$.

evidence. All three measures d, r, and l satisfy the fundamental Peirceian desideratum $\mathcal{A}$ (and $\mathcal{A}'$). However, measures s and $\mathfrak{r}$ would seem to be inadequate in their handling of independent evidence, even at this most basic level.

Unfortunately, the agreement between d, r, and l ends here. In the next section, I will describe a symmetry desideratum which is satisfied by d and l (and s and $\mathfrak{r}$), but violated by r. This will narrow down the field further to two measures (d and l) which seem to cope adequately (at a very basic level) with independent evidence. Later, I will propose additional, probabilistic constraints on accounts of independent evidence that narrow the field even more.

3.2.2 A Negation Symmetry Desideratum

If two pieces of evidence are confirmationally independent regarding H, then they should also be confirmationally independent regarding $\bar{H}$. Negation symmetry in the confirmational independence relation seems highly intuitive.[53] After all, if the degree to which E_1 confirms H doesn't depend on whether E_2 is already known, then why should the degree to which E_1 confirms $\bar{H}$ depend on whether E_2 is already known? In our confirmation theoretic framework, this intuitive negation symmetry principle gets formalized as follows:

$$(\mathcal{S})\ \text{If}\ \mathfrak{c}(H, E_1 \mid E_2) = \mathfrak{c}(H, E_1)\ \text{and}\ \mathfrak{c}(H, E_2 \mid E_1) = \mathfrak{c}(H, E_2),\ \text{then}\ \mathfrak{c}(\bar{H}, E_1 \mid E_2) = \mathfrak{c}(\bar{H}, E_1)\ \text{and}\ \mathfrak{c}(\bar{H}, E_2 \mid E_1) = \mathfrak{c}(\bar{H}, E_2).$$

The following theorem states that each of our five Bayesian relevance measures — *except* r — satisfies $\mathcal{S}$.

Theorem 7. *Each of the measures d, l, s, and $\mathfrak{r}$ satisfies $\mathcal{S}$, but r violates $\mathcal{S}$.*[54]

[53] Many varieties of independence satisfy this kind of negation symmetry requirement (*e.g.*, both logical independence and probabilistic independence are negation-symmetric).

The two high-level[55] desiderata $\mathcal{A}$ and $\mathcal{S}$ narrow the field of four relevance measures down to two (d and l) which seem — so far — to explicate the concept of independent evidence.[56] Below, I will propose a low-level, probabilistic constraint that rules out the difference measure d and all other relevance measures, except those ordinally equivalent to the log-likelihood ratio measure l. But, before we get to that, I'd like to consider an interesting (albeit special) class of cases in which the log-likelihood-ratio measure l seems to give more intuitive results than the difference measure d. This should help pave the way for subsequent arguments.

3.2.3 Conclusive Evidence and Measures of Confirmation

So far, we have seen two sets of simple and intuitive, high-level desiderata each of which narrows the field of five candidate measures down to two: d and l. In the sections below, I will describe a set of low-level, probabilistic constraints that provide an ultimate adjudication between d and l (in favor of l). But, before I present that material, I'd like to talk briefly about cases in which E is *conclusive* for H (relative to K). I think that by looking at this special case, one can begin to see some advantages l has over d. Moreover, I think that this class of cases is one in which Good's requirement of "independence of the prior probability of H" makes sense (and can be made precise and non-superficial).

We say that E provides *conclusive support or confirmation* for H (relative to K)

---

[54]This theorem is closely related to a result reported in Eells and Fitelson (2001) which says that each of our five Bayesian relevance measures — *except* r — satisfies the following *hypothesis symmetry* condition: (HS) $\mathfrak{c}(H, E \mid K) = -\mathfrak{c}(\bar{H}, E \mid K)$. Note that (HS) is Carnap's (14) [§3.1.2]. See, Eells and Fitelson (2001) for some reasons to think that an adequate measure of support $\mathfrak{c}$ *should* satisfy (HS). It is interesting to note that (HS) entails our negation symmetry desideratum $\mathcal{S}$, but not conversely (proof omitted). So, $\mathcal{S}$ is a *strictly weaker* desideratum than (HS).

[55]When I call a set of constraints "*high-level* desiderata," I mean that the constraints can be stated entirely at the level of the measure $\mathfrak{c}$ — without having to appeal to any properties of the underlying (*low-level*) probability function Pr in terms of which the measures $\mathfrak{c}$ are defined.

[56]See Eells and Fitelson (2001) for an independent set of high-level desiderata which also narrow the field to the two measures d and l. *Pace* Milne (1996), d and l seem, in many ways, to be the two most serious candidates for "the one true measure of confirmation."

just in case $E \& K \vDash H$. I take it as intuitively clear that the strength of the support E provides for H in this case should *not* depend on how probable H is (*a priori*). Just think about deductively valid arguments for a moment. Do we want to say that the strength of a valid argument should depend on how confident we were in the conclusion of the argument before we thought about the premises? Clearly, the answer to this question is "no." After all, evidential support is supposed to be a measure of how strong the evidential *relationship* between E and H is, and deductive entailment is the strongest that such a relationship can possibly get. If E is *conclusive* for H, then H's *a priori* probability should, intuitively, be *irrelevant* to how strong the (maximal, deductive) evidential relationship between E and H is. It seems to me that this simple idea can be translated into the following intuitive desideratum for adequate measures of degree of support $\mathfrak{c}(H, E \mid K)$:

($\mathcal{K}$) If E provides conclusive support or confirmation for H (relative to K), then $\mathfrak{c}(H, E \mid K)$ should be *maximal* (*viz.*, *constant*), and should not depend on the prior probability of H [$\Pr(H \mid K)$].

Interestingly, the only measure (among our five candidates) that satisfies $\mathcal{K}$ is l. To see why l satisfies $\mathcal{K}$, it is easier to work with the following, ordinally equivalent[57] measure l^*, which was proposed and defended by Kemeny and Oppenheim (1952).[58]

$$l^*(H, E \mid K) = \frac{\Pr(E \mid H \,\&\, K) - \Pr(E \mid \bar{H} \,\&\, K)}{\Pr(E \mid H \,\&\, K) + \Pr(E \mid \bar{H} \,\&\, K)}$$

It is easy to show that l^* has the following property (proof omitted), which explains why l^* satisfies our conclusive evidence desideratum $\mathcal{K}$.[59]

[57]Measures l and l^* are ordinally equivalent since l^* is an isotone function of l [$l^* = \sinh(l/2)$].

[58]Kemeny and Oppenheim (1952) provide an interesting (and deep) desideratum/explicatum argument for l^*. I am currently working on an analysis of their argument in favor of l^*.

[59]Strictly speaking, $E \,\&\, K \nvDash H$ and $E \,\&\, K \nvDash \bar{H}$ imply that $l^*(H, E \mid K) \in (-1, 1)$ *only if* it assumed that the probability spaces are *finite*, or if Pr is assumed to be *regular* (a regular probability function assigns probability zero *only* to $\bot$ and probability one *only* to $\top$).

$$(\mathcal{K}') \qquad l^*(H, E \mid K) \begin{cases} = 1 & \text{if } E \mathbin{\&} K \vDash H, \\ \in (-1, 1) & \text{if } E \mathbin{\&} K \nvDash H \text{ and } E \mathbin{\&} K \nvDash \bar{H}, \\ = -1 & \text{if } E \mathbin{\&} K \vDash \bar{H}. \end{cases}$$

Intuitively, $\mathcal{K}'$ is exactly the kind of property a measure of support should have. Any measure satisfying $\mathcal{K}'$ will take on its *maximal* (and *constant*) value when E is *conclusive* for H, its *minimal* (and *constant*) value when E is conclusive for the *denial* of H, and an intermediate value when E is *deductively independent* of H.

Perhaps this is the class of examples Good (1984) had in mind when he insisted that $\mathfrak{c}(H, E \mid K)$ should not "depend on the prior probability of H." In this class of cases, it is both true and intuitive that l^* (hence, by ordinal equivalence, l) does not depend on the prior probability of H. In any case, we should probably not put too much stock on deductive cases of the kind discussed in this section. This section was mainly intended as an "intuition pump" to prime the reader for the sections below. Below, I will describe a much more interesting, *inductive* variety of independence which will also single out l as a superior measure of confirmation.

3.2.4 Screening-Off and Confirmational Independence

3.2.4.1 Wittgenstein's Example and Sober's Probabilistic Analysis

Wittgenstein (1953) 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 "double check." This does not seem to be an effective strategy for corroboration. To fix our ideas, let's assume that a story in the New York Times (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 story in the) same issue of the NYT. Intuitively, the degree to which the conjunction $E_1 \mathbin{\&} E_2$ confirms H is no greater than the degree to which E_1 *alone* confirms H. Also, it seems intuitive that an *independent* report E' (say, one heard on a NPR broadcast) *would* serve to corroborate the NYT story. How can we explain the epistemic difference between these two cases? Intuitively, a NYT report (E_i) and a NPR report (E') provide *independent* support for H in a way that two NYT reports (E_1, E_2) do not.

Sober (1989) offers an illuminating and suggestive probabilistic analysis of this problem. Sober explains that the probabilistic structure of this example is a *conjunctive fork*, in which E_i and E' are joint effects of a common cause H. Sober also points out [as Reichenbach (1956, page 159) first did] that E_i and E' will *not* be *un*conditionally probabilistically independent in such a case. So, it *can't* be probabilistic independence of the evidence *simpliciter* which is responsible for our intuitive judgment that E_i and E' are confirmationally independent *regarding* H in this example. Is there *some* probabilistic feature of this example which undergirds our intuition? It seems to me (as it did to Sober) that the relevant point is that [in the terminology of Reichenbach (1956, page 189)] H *screens-off* E_i from E'. That is, it is the fact that E_i and E' are probabilistically independent *conditional on the hypothesis* H (and its denial) that explains our intuition that E_i and E' are *confirmationally independent regarding* H. To appreciate Sober's explanation, it helps to picture the probabilistic (causal) structure of the example, as in Figure 1:

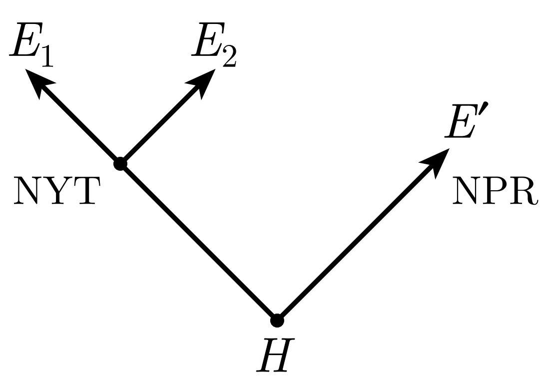

Figure 1: Picturing the structure of the Wittgenstein/Sober example

As the picture makes clear, H does *not* screen-off E_1 from E_2 (although, perhaps the state of the NYT printing press prior to printing *does* screen them off). 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*.

Sober's analysis of Wittgenstein's example provides informal motivation for the following two central points concerning the nature of confirmational independence and its intuitive relation to probabilistic screening-off:

- Confirmational independence is inherently a *three-place* relation. That is, when we say E_1 and E_2 are *confirmationally independent regarding* H, we are *not* saying that E_1 and E_2 are *un*conditionally independent of *each other*. We are talking about a kind of (*ternary*) independence relation that depends crucially on the hypothesis H.

- *Screening-off* of E_1 from E_2 by H is (intuitively) intimately connected with confirmational independence of E_1 and E_2 regarding H.

In the next section, I will describe a more general and formal probabilistic model that is intended to make the connection between screening-off and confirmational independence more precise. This formal model will also allow us to generate concrete, numerical examples which will, ultimately, be used to show that only the log-likelihood ratio measure l properly handles the (general) relationship between probabilistic screening-off and confirmational independence.

3.2.4.2 A Formal Model

To formally motivate the general connection between probabilistic screening-off and confirmational independence, I will use a simple, abstract model. I will call this model the *urn model*.[60] The background evidence K for the urn model is assumed at the outset to consist of the following information:

> An urn has been selected at random from a collection of urns. Each urn contains some balls. In some of the urns the proportion of white balls to other balls is x and in all the other urns the proportion of white balls is y, $0 < x, y < 1$. The proportion of urns of the first type is z, $0 < z < 1$. Balls are to be drawn randomly from the selected urn, with replacement.

Let H be the hypothesis that the proportion of white balls in the urn is x. Let W_i state that the ball drawn on the i^{th} draw ($i \geq 1$) is white. I take it as intuitively clear that W_1 and W_2 are mutually confirmationally independent regarding H, regardless of the values of x, y, and z. Hence, I propose the following adequacy condition for measures of degree of confirmation:

(UC) If $\mathfrak{c}$ is an adequate measure of degree of confirmation then, both $\mathfrak{c}(H, W_1 \mid W_2) = \mathfrak{c}(H, W_1)$, and $\mathfrak{c}(H, W_2 \mid W_1) = \mathfrak{c}(H, W_2)$ for all urn examples, regardless of the values of x, y, and z.[61]

What probabilistic feature of the urn model could be responsible for the (presumed) fact that W_1 and W_2 are confirmationally independent regarding H? The feature cannot depend on the *values* of the probabilities involved, since we did not specify what these are except to say that they are not zero or one (a requirement

[60] The urn model is due to Patrick Maher.

[61] Ellery Eells (private communication) worries that for extreme (or near extreme) values of x, y, or z, this intuition might break down. He may be right about this (although, as a defender of l, I will insist that any such breakdown can be explained away, and is probably just a psychological "edge effect," owing to the extremity of the values of x, y or z, and not to considerations relevant to their confirmational independence *per se*). However, in the Appendix (Theorem 8), I show that the measures d, r, s, and $\mathfrak{r}$ *fail* to obey this intuition, even in cases where the values of x, y, and z are all *arbitrarily far from extreme*. As a result, d, r, s, and $\mathfrak{r}$ will not even judge E_1 and E_2 as confirmationally independent regarding H in *Sober's* example. This seems unintuitive, and should cast doubt on the adequacy of d, r, s, and $\mathfrak{r}$.

imposed to ensure that the relevant conditional probabilities are all defined). Moreover, as we saw in Sober's example, the feature cannot depend on the *un*conditional probabilistic independence of W_1 and W_2, since W_1 and W_2 will *not*, in general, be independent of *each other* (*e.g.*, if each of W_1 and W_2 *individually* confirms H). This does not leave much. Two considerations that remain are that the following two identities hold in all urn examples:

(18) $$\Pr(W_1 \,\&\, W_2 \mid H) = \Pr(W_1 \mid H) \cdot \Pr(W_2 \mid H)$$

(19) $$\Pr(W_1 \,\&\, W_2 \mid \bar{H}) = \Pr(W_1 \mid \bar{H}) \cdot \Pr(W_2 \mid \bar{H})$$

Identities (18) and (19) state that H *screens-off* W_1 from W_2 (or, equivalently, W_2 from W_1). What I am suggesting, then, is that screening-off by H of W_1 from W_2 is a *sufficient* condition for W_1 and W_2 to be mutually confirmationally independent regarding H. This suggests that (UC) might be strengthened to the following screening-off adequacy condition for measures of confirmation:

(SC) If $\mathfrak{c}$ is an adequate measure of confirmation, and if H screens-off E_1 from E_2, then $\mathfrak{c}(H, E_1 \mid E_2) = \mathfrak{c}(H, E_1)$ and $\mathfrak{c}(H, E_2 \mid E_1) = \mathfrak{c}(H, E_2)$.

I find (SC) an attractive principle; but, for the purposes of this paper, I will use only the weaker (and perhaps more intuitive) adequacy condition (UC).[62] The following theorem states that the only measure among our four measures d, r, l, and s that satisfies (UC) is the log-likelihood ratio measure l:

Theorem 8. *The measures d, r, s, and $\mathfrak{r}$ violate* (UC)*, but l satisfies* (UC)*.*[63]

[62] Heckerman (1988, page 19) has suggested an adequacy condition that is equivalent to (SC). He gives no justification for this principle. I take the urn model to be a *partial* justification of (SC). However, I prefer the present approach since it makes use only of the weaker (and, I think, more intuitive) (UC). Incidentally, I do *not* think that screening-off by H is a *necessary* condition for mutual confirmational independence regarding H (neither does Heckerman). I discuss this issue further in the Appendix, when I prove Theorem 8 (see §A.8 for a counterexample).

Thus, only the log-likelihood ratio l satisfies the low-level, probabilistic screening-off desideratum. I think this is a compelling reason to favor the log-likelihood ratio measure over the other measures currently defended in the philosophical literature (at least, when it comes to judgments of confirmational independence regarding a hypothesis).[64] As such, this provides a possible (at least, partial) solution to the problem of the plurality of Bayesian measures of confirmation described in Fitelson (1999). In the next section, I will discuss another application of my account of independent evidence.

3.3 An Application to Evidential Diversity

Philosophers of science dating back at least to Carnap (1945) have shared the intuition that collections of evidence that are 'diverse' or 'varied' should (*ceteris paribus*[65]) confirm more strongly than collections of evidence that are 'narrow' or 'homogeneous'. I have elsewhere [see Fitelson (1996)] called this the *confirmational significance of evidential diversity* (CSED). I suspect that the notion of *independent* evidence can undergird, at least partially, (some of) our intuitions about the

[63]Heckerman (1988) claims to prove a much more ambitious, and closely-related result. He claims to show that only measures that are ordinally equivalent to l satisfy (SC). Unfortunately, his argument is fallacious for subtle mathematical reasons — see Halpern (1999a, 1999b). In particular, like the arguments of Milne (1996) and Good (1984), Heckerman's argument presupposes that an agent's probability space is infinite, and satisfies some rather strong (unmotivated) mathematical constraints [see Halpern (1999a, 1999b)]. Unlike Heckerman's argument, my argument makes use only of the finitistic adequacy condition (UC), and requires no additional, strong mathematical presuppositions.

[64]The intimate connection between probabilistic screening-off of the kind described here and our intuitive judgments of independent inductive support has been pointed out by several recent authors (and used by some a reason to favor likelihood-ratio based measures of support), including: Good (1983), Pearl (1988), Heckerman (1988), and Schum (1994).

[65]See Fitelson (1996) and §3.3.2 below for an elaboration of the *ceteris paribus* conditions that are tacitly presupposed in the Bayesian explication of CSED offered by Horwich (1982). I will later discuss the *ceteris paribus* clauses implicit in Howson and Urbach's (1993) 'correlation' approach to CSED. Carnap's original (1945, page 94) explication of CSED also requires some rather sophisticated *ceteris paribus* conditions. But, since Carnap's original account of CSED does not make explicit use of d, r, l, s, or $\mathfrak{r}$, it is beyond the scope of this monograph.

significance of *diverse* evidence. At least one recent philosopher of science seems to share this suspicion. Sober (1989) shows (essentially[66]) that the log-likelihood ratio measure l satisfies the following high-level diversity desideratum:

($\mathcal{D}$) If each of E_1 and E_2 individually confirms H, and if E_1 and E_2 are confirmationally independent regarding H according to $\mathfrak{c}$, then $\mathfrak{c}(H \mid E_1 \,\&\, E_2) > \mathfrak{c}(H \mid E_1)$ and $\mathfrak{c}(H \mid E_1 \,\&\, E_2) > \mathfrak{c}(H \mid E_2)$.[67]

It is a direct corollary of Theorem 6 that — according to *all three* measures of confirmation d, r, and l — two pieces of *independent* confirmatory evidence will always provide stronger confirmation that either one of them provides individually. In other words, we have already shown that the three most popular measures of confirmation d, r, and l *all* satisfy $\mathcal{D}$.[68] It seems to me that $\mathcal{D}$ could be used to provide a rather simple and elegant (partial[69]) Bayesian account of CSED. The basic idea behind such an approach would be that it is not evidence of different 'kinds' *per se* that will boost confirmational power. Rather, it is *data*

[66]Strictly speaking, Sober proves something *weaker* than this. He proves that l satisfies the consequent of $\mathcal{D}$ under the *stronger* (wrt l) assumption that H *screens-off* E_1 from E_2. Our result is also more general than Sober's in the sense that it applies not only to l but to d and r as well (*i.e.*, our result $\mathcal{D}$ is not as sensitive to the choice of measure of confirmation).

[67]The following, low-level diversity desideratum is also of interest:

($\mathcal{D}'$) If each of E_1 and E_2 individually confirms H, and if H screens-off E_1 from E_2, then $\mathfrak{c}(H, E_1 \,\&\, E_2) > \mathfrak{c}(H, E_1)$ and $\mathfrak{c}(H, E_1 \,\&\, E_2) > \mathfrak{c}(H, E_2)$.

We know from Theorems 6 and 8 that l satisfies $\mathcal{D}'$. And, we have recently discovered that measures s and $\mathfrak{r}$ *violate* D' (countermodels omitted). It remains open whether d or r satisfy D' (computer searches indicate that they "probably" do, but no proofs have been found). See §4.1.

[68]Although s and $\mathfrak{r}$ violate $\mathcal{A}$, it remains open whether s or $\mathfrak{r}$ violates $\mathcal{D}$. If s and $\mathfrak{r}$ satisfy $\mathcal{D}$, then this would make $\mathcal{D}$ *totally insensitive to choice of measure* (at least, among d, r, l, s, and $\mathfrak{r}$). See §4.1 for further discussion of the remaining open questions concerning $\mathcal{D}$ and $\mathcal{D}'$.

[69]I do not mean to suggest that confirmational independence can be used to undergird *all* of our intuitions about the value of diverse evidence. But, I do think that there are many important scientific cases that fit this mold. For instance, the intuition that evidence from independent domains of application (*e.g.*, celestial *vs* terrestrial domains) of a theory often confirm more strongly than the same amount of evidence from domains of application that are not independent is a canonical example of the kind of intuition I have in mind here. Moreover, Sober (1989, page 124) explains how the notion of independent evidence regarding a hypothesis can be useful in the context of phylogenetic inference (*e.g.*, the problem of inferring the character states of ancestors from the observed character states of their descendants).

whose confirmational power is maximal, given the evidence we already have that are confirmationally advantageous. And, $\mathcal{D}$ provides a robust, general sufficient[70] condition for this sort of confirmational boost.

It is *not* generally the case (as was pointed out by Carnap (1962)) that two pieces of confirmatory evidence *simpliciter* will always provide stronger confirmation than just one. With $\mathcal{D}$, we have identified a very general sufficient condition for increased confirmational power. The slogan behind $\mathcal{D}$ might be "Two pieces of *independent* evidence are better than one." One nice thing about $\mathcal{D}$ is that *it does not depend sensitively on one's choice of measure of confirmation*. Below, I compare the present approach to CSED with a recent Bayesian alternative proposed by Howson and Urbach (1993).

3.3.1 Comparison with the 'Correlation' Approach

Howson and Urbach (1993) propose a different way to account for our intuitions about CSED.[71] This approach asks us to consider *not* whether E_1 and E_2 are confirmationally independent *regarding* H. Rather, Howson and Urbach (1993) suggest that the important thing is whether or not E_1 and E_2 are *un*conditionally stochastically independent. Howson and Urbach (1993, pages 113–114, my italics) summarize the their 'correlation' account as follows:

> Evidence that is varied is often regarded as offering better support to a hypothesis than an equally extensive volume of homogeneous evidence . . . According to the Bayesian, if two data sets are entailed by a hypothesis (or

[70] As was the case with (UC) and (SC), I am *not* claiming that $\mathcal{D}$ is a *necessary* condition for increased confirmational power in this sense (indeed, I suspect it is *not* — see §A.8).

[71] Earman (1992) discusses a similar approach. Basically, many of the same criticisms will apply to his account. I will focus on the account of Howson and Urbach, since their characterization of the 'correlation' approach is closer in spirit to my presentation. See Forster (1995) for some independent criticisms of Earman's particular 'correlation' explication of CSED.

> have similar probabilities relative to it[72]), and one of them confirms more strongly than the other, this must be due to a corresponding difference between the data in their probabilities ... The idea of similarity between items of evidence is expressed naturally in probabilistic terms by saying that e_1 and e_2 are similar if $P(e_2 \mid e_1)$ is higher than $P(e_2)$, and one might add that the more the first probability exceeds the second, the greater the similarity. *This means that e_2 would provide less support if e_1 had already been cited as evidence than if it was cited by itself.*

The most charitable interpretation of the above proposal of Howson and Urbach would seem to be the following rather complicated nested conditional:

($\mathcal{H}$) *If* the following probabilistic '*ceteris paribus* clause' is satisfied:

(CP) $\Pr(E_1 \mid H) = \Pr(E_2 \mid H) = \Pr(E_1 \,\&\, E_2 \mid H) = 1$,

then if $\Pr(E_2 \mid E_1) > \Pr(E_2)$, then $\mathfrak{c}(H, E_2 \mid E_1) < \mathfrak{c}(H, E_2)$.[73]

In other words, Howson and Urbach are claiming that (*ceteris paribus*[74]) pieces of evidence E_1 and E_2 that are *un*conditionally positively correlated will *not* be confirmationally independent regarding a hypothesis H (and, that E_1 and E_2 will tend to *cancel* each other's support for H in such cases). I see several serious problems with Howson and Urbach's proposal $\mathcal{H}$.[75]

As we have already seen in Sober's conjunctive fork example, pieces of *confirmationally independent* evidence will often be *un*conditionally positively correlated

[72] Howson and Urbach's parenthetical remark that their '*ceteris paribus* condition' (CP) can be weakened to (CP') $\Pr(E_1 \mid H) = \Pr(E_2 \mid H)$ — while preserving the general truth of the main tenet ($\mathcal{H}$) of their account of CSED — is *false*. See footnote 108 in the Appendix (in the proof of Theorem 9) for a proof that this parenthetical remark is incorrect.

[73] In fact, Howson and Urbach seem to be making an even stronger, *quantitative* claim. They seem to be saying that if (CP) is satisfied, then the great*er* $\Pr(E_2 \mid E_1)$ is than $\Pr(E_2)$, the less*er* $\mathfrak{c}(H, E_2 \mid E_1)$ will be than $\mathfrak{c}(H, E_2)$. I have chosen to criticize the (weaker) *qualitative* interpretation $\mathcal{H}$, since $\mathcal{H}$'s falsity entails the falsity of the stronger, quantitative claim.

[74] Howson and Urbach's (CP) is slightly stronger than the probabilistic *ceteris paribus* clause that is needed to shore-up Horwich's (1982) account of CSED [see Fitelson (1996) and §3.3.2].

[75] Note that Howson and Urbach's $\mathcal{H}$ only purports to explain why a *lack* of evidential 'diversity' can be *bad*. $\mathcal{H}$ cannot tell us why or how evidential 'diversity' can be *good*.

(and often *strongly* so). Newspaper reports (E_1) and radio reports (E_2) about the outcome (H) of a baseball game often fail to be *un*conditionally independent. This (in and of itself) does nothing to undermine our intuition that E_1 and E_2 are *confirmationally independent* regarding H. Moreover, this example is representative of a wide range of cases. The conjunctive fork structure is common in (intuitive) examples of confirmational independence. For example, consider what doctors do when they seek independent confirmation of a diagnosis. They look for confirmationally independent corroborating symptoms. Such symptoms will typically be unconditionally correlated with already observed symptoms. But, *conditional on the relevant diagnostic hypothesis*, confirmationally independent symptoms will tend to be stochastically independent. It is *conditional* independence that is relevant here, not unconditional independence.

At best, Howson and Urbach have shown (*via* $\mathcal{H}$) that confirmational independence and unconditional stochastic dependence cannot co-occur in the extreme, deductive cases in which (CP) holds.[76] If $\mathcal{H}$ were true for an interesting class of Bayesian confirmation measures $\mathfrak{c}$, then Howson and Urbach's account would, at least, provide some useful information about the relationship between confirmational independence and unconditional stochastic independence in the case of deductive evidence. Unfortunately, as the following theorem states, among the five measures we have considered, Howson and Urbach's $\mathcal{H}$ is satisfied *only* by the log-ratio measure r and Carnap's measure $\mathfrak{r}$, both of which we have already shown to be inadequate in several important respects.[77]

Theorem 9. *$\mathcal{H}$ is true if $\mathfrak{c} = r$ or $\mathfrak{c} = \mathfrak{r}$, but $\mathcal{H}$ is false if $\mathfrak{c} = d$, $\mathfrak{c} = l$, or $\mathfrak{c} = s$.*

[76] The fact that odd things can happen in such extreme cases was pointed out by Sober (1989, page 279). There, Sober explains that many of the salient epistemological differences between independent and dependent evidence collapse in the extreme (deterministic) case.

[77] Myrvold (1996) has independently articulated some of these same criticisms of Howson and Urbach's account. Moreover, Myrvold nicely shows how to remedy many of these problems, by being sensitive to *conditional* independence (as well as *un*conditional independence). While this is certainly a step in the right direction, the new 'correlation' account presented in Myrvold (1996) is strongly sensitive to an inadequate choice (r) of measure of confirmation (proof omitted).

Howson and Urbach must either embrace the unattractive option of defending their chosen measure r, or they must defend some other measure of confirmation which satisfies $\mathcal{H}$ (*e.g.*, Carnap's measure $\mathfrak{r}$, which we have also shown to be inadequate in several ways). In either case, Howson and Urbach must reject the general connection (SC) between screening-off and confirmational independence, since (SC) and $\mathcal{H}$ are logically incompatible. That is, in the case of deterministic conjunctive forks, (SC) and $\mathcal{H}$ cannot both be true.[78]

3.3.2 Comparison with Wayne and Horwich on CSED[79]

Wayne (1995) gives one reconstruction of Horwich's (1982) Bayesian account of the value of evidential diversity. He then shows that there are counterexamples to this reconstruction of Horwich's explication of CSED. Such counterexamples would undermine Horwich's account of CSED, *if* Wayne's reconstruction were a charitable one. Presently, I argue that Wayne's reconstruction of Horwich's account of CSED is uncharitable. As a result, his criticisms are not genuine problems for Horwich. This does *not* mean that Horwich's explication of CSED — charitably reconstructed — is unproblematic. On the contrary, after my analysis of Wayne's critique, I discuss several remaining problems for Horwich's account. In the end, I conclude that Horwich's Bayesian explication of CSED is inadequate.

3.3.2.1 Wayne's Reconstruction of Horwich's Account

In a typical confirmation theoretic context $\mathcal{C}$, we have a hypothesis under test H_1 and $n-1$ competing hypotheses $H_2, \ldots, H_n$, where the n hypotheses are assumed to be mutually exclusive and exhaustive. Wayne's (1995) reconstruction

[78] This is easily proved. Assume that (CP) obtains (which implies that H screens-off E_1 from E_2). In such a case, H, E_1, and E_2 will form a (deterministic) conjunctive fork. Now, if $\mathcal{H}$ is true, then $\mathfrak{c}(H, E_2 \mid E_1) < \mathfrak{c}(H, E_2)$. But, if (SC) is true, then $\mathfrak{c}(H, E_2 \mid E_1) = \mathfrak{c}(H, E_2)$. Therefore, in the case of deterministic conjunctive forks, (SC) and $\mathcal{H}$ cannot both be true. ❑

[79] Much of the material in this section appears in Fitelson (1996).

of Horwich's (1982) explication of CSED involves the following three propositions concerning such contexts:

($\mathcal{H}_1$) One collection of evidence E_1 is more *confirmationally diverse* (c-diverse) than another collection of evidence E_2 in context $\mathcal{C}$ iff $(\forall i \neq 1)[\Pr(E_1 \mid H_i \,\&\, K_{\mathcal{C}}) < \Pr(E_2 \mid H_i \,\&\, K_{\mathcal{C}})]$.[80]

The intuition behind $\mathcal{H}_1$ is that the more c-diverse collection of evidence is supposed to "rule-out most plausible alternatives" to the hypothesis under test. It is for this reason that Horwich's account has been called 'eliminativist'.[81]

($\mathcal{H}_2$) E_1 confirms H more strongly than E_2 confirms H if and only if $r(H, E_1) > r(H, E_2)$, where the *ratio* measure of degree of confirmation $r(H, E)$ is defined as follows: $r(H, E) =_{df} \frac{\Pr(H \mid E)}{\Pr(H)}$.

($\mathcal{H}_3$) For every confirmational context $\mathcal{C}$, if E_1 is more c-diverse than E_2 in $\mathcal{C}$, then E_1 confirms H_1 (*i.e.*, the hypothesis under test in $\mathcal{C}$) more strongly than E_2 confirms H_1 in $\mathcal{C}$.

According to Wayne (1995), $\mathcal{H}_3$ captures the kernel of Horwich's account of CSED. In the next section, we will look at a counterexample to $\mathcal{H}_3$ due to Wayne (1995).

3.3.2.2 Wayne's Counterexample to $\mathcal{H}_3$

Wayne (1995, page 119) asks us to:

[80] Where, the proposition $K_{\mathcal{C}}$ encodes the *background knowledge* in confirmational context $\mathcal{C}$. Hereafter, I will, for simplicity's sake, drop explicit reference to $K_{\mathcal{C}}$ in probability statements. It is to be understood, of course, that we are uniformly conditioning Pr on $K_{\mathcal{C}}$, whenever we make a Bayesian confirmational comparison.

[81] Notice that $\mathcal{H}_1$ is sufficient but not necessary for E_1's "ruling-out" *most* alternatives to H_1 in $\mathcal{C}$. As we'll see, this added strength is needed to shore-up Horwich's account of CSED. See the Appendix §A.11–§A.13 for all technical details pertaining to this section of the monograph.

> … consider a simple context $\mathcal{C}_w$ in which only three hypotheses have substantial prior probabilities, $\Pr(H_1) = 0.2$, $\Pr(H_2) = 0.2$, $\Pr(H_3) = 0.6$, and two data sets E_1 and E_2 such that:
>
> $$\begin{array}{ll} \Pr(E_1 \,|\, H_1) = 0.2 & \Pr(E_2 \,|\, H_1) = 0.6 \\ \Pr(E_1 \,|\, H_2) = 0.4 & \Pr(E_2 \,|\, H_2) = 0.5 \\ \Pr(E_1 \,|\, H_3) = 0.4 & \Pr(E_2 \,|\, H_3) = 0.6 \end{array}$$
>
> This is plainly a paradigm case of $\mathcal{H}_1$: for all H_i, $\Pr(E_1 \,|\, H_i)$ is significantly less than $\Pr(E_2 \,|\, H_i)$. Yet, a straightforward substitution shows that $\mathcal{H}_3$ is violated! Thus, we obtain the counterintuitive result that the *similar* evidence lends a greater boost to the hypothesis under test than does the *diverse* evidence … Horwich's account fails to reproduce our most basic intuition about diverse evidence.[82]

Wayne is right about $\mathcal{C}_w$ in the following two respects.

(20) In $\mathcal{C}_w$, E_1 is more c-diverse than E_2.

(21) In $\mathcal{C}_w$, E_2 confirms H_1 more strongly than E_1 confirms H_1, according to the ratio measure r.

Hence, $\mathcal{C}_w$ *is* a legitimate counterexample to $\mathcal{H}_3$. In the next section, I will discuss some aspects of Wayne's example that he neglects to mention. Then, I will reflect on what the existence of this counterexample implies — and *doesn't* imply — about Horwich's account of CSED.

3.3.2.3 Why Wayne's Counterexample is Not Salient

Here is a fact about Wayne's counterexample to $\mathcal{H}_3$ that he neglects to mention.

(22) In $\mathcal{C}_w$, E_2 confirms H_1; whereas, E_1 *dis*confirms H_1.

[82] I have taken the liberty of translating this passage from Wayne (1995) into my notation.

Wayne has certainly described a confirmational context $\mathcal{C}_w$ in which a less c-diverse data set confirms the hypothesis under test more strongly than a more c-diverse data set does. But, as it turns out, $\mathcal{C}_w$ is also a context in which the more c-diverse evidence *dis*confirms the hypothesis under test; whereas, the less c-diverse evidence confirms the hypothesis under test. What does this mean?

3.3.2.4 Charitably Reconstructing Horwich's Account

As far as I can tell, (22) shows that $\mathcal{H}_3$ must *not* be what Horwich has in mind in his explication of CSED. Surely, Horwich would *not* want to say that more c-diverse *dis*confirmatory evidence should confirm more strongly than less c-diverse confirmatory evidence. To say the least, this would not be in the spirit of the Bayesian definition of confirmation.

A more charitable reconstruction of Horwich's account of CSED should add a suitable probabilistic *ceteris paribus* clause to $\mathcal{H}_3$. In such a reconstruction, Wayne's $\mathcal{H}_3$ might be replaced by:

($\mathcal{H}_3'$) If CP then $\mathcal{H}_3$.

Where CP is an appropriate probabilistic ceteris paribus clause. Wayne's counterexample teaches us that, at the very least, CP should entail:

(CP_1) Both E_1 and E_2 confirm H_1 in $\mathcal{C}$.

Indeed, CP_1 would avoid the counterexample raised by Wayne. Moreover, it would insure that $\mathcal{H}_3'$ does not contradict the Bayesian definition of confirmation (as Wayne's $\mathcal{H}_3$ does).

Interestingly, CP_1 is *not* a sufficient ceteris paribus clause. For, CP_1 does not entail $\mathcal{H}_3$. We will need to make CP substantially stronger than CP_1 in order to make $\mathcal{H}_3'$ a theorem of the mathematical theory of probability. There are many ways to define sufficient ceteris paribus clauses in this sense.[83] Here is one such proposal that I think remains faithful to what Horwich has in mind:

$(CP*)$ CP_1, *and* $\Pr(E_1 \mid H_1) = \Pr(E_2 \mid H_1)$ in $\mathcal{C}$.

$CP*$ says that E_1 and E_2 both confirm H_1 in $\mathcal{C}$, *and* that E_1 and E_2 are '$\mathcal{C}$-commensurate', in the sense that the hypothesis under test has the same likelihood (*i.e.*, goodness of fit) with respect to both E_1 and E_2 in $\mathcal{C}$. This ceteris paribus clause seems to be implicit in Horwich's depiction of the kinds of confirmational contexts he has in mind. Figure 2 shows the kind of confirmational contexts and comparisons that Horwich (1982, pages 119–120) uses as canonical illustrations of his account of CSED.

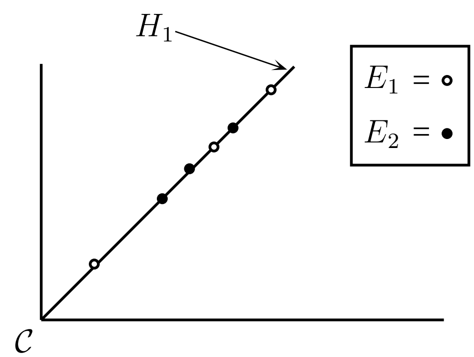

Figure 2: A canonical Horwichian example of CSED

Figure 2 depicts a canonical confirmation theoretic context $\mathcal{C}$ in which the hypothesis under test H_1 fits two data sets E_1 and E_2 equally well, in accordance

[83] Hellman (1997) proposes the following alternative sufficient ceteris paribus clause:

$(CP\dagger)$ CP_1, *and* $\Pr(E_1 \mid H_1) - \Pr(E_1) = \Pr(E_2 \mid H_1) - \Pr(E_2)$ in $\mathcal{C}$.

It is true that $CP\dagger$ is sufficient for $\mathcal{H}_3$. However, $CP\dagger$ is clearly *not* the kind of Bayesian proposal that Horwich (1982) has in mind. In Horwich's canonical examples, it is typically assumed that $\Pr(E_1 \mid H_1) = \Pr(E_2 \mid H_1)$ (see below for more on this point). Moreover, Horwich wants sets of evidence with *greater* c-diversity to have *lesser* prior probability (*e.g.*, Horwich wants $\Pr(E_1) < \Pr(E_2)$ in his canonical example). These two constraints jointly entail that $CP\dagger$ does *not* hold. So, while Hellman's alternative makes sense from a generic Bayesian point of view, it is not a faithful reconstruction of *Horwich*'s Bayesian explication of CSED. See Kruse (1999) and Maher (1997) for further (and deeper) criticisms of Hellman (1997).

with $CP*$. Moreover, E_1 is more *intuitively* diverse (i-diverse) than E_2, since the abscissa values of E_1 are more spread-out than the abscissa values of E_2.[84] Horwich (1982) seems only to be claiming that — other things being equal (those other things being the *likelihoods* $\Pr(E_1 \mid H_1)$ and $\Pr(E_2 \mid H_1)$) — more diverse[85] sets of evidence (*e.g.*, E_1) will confirm the hypothesis under test (*e.g.*, H_1) more strongly than less diverse sets of evidence (*e.g.*, E_2) will. This is an appropriate time to state the following theorem:

($\mathcal{H}_3*$) If $CP*$, then $\mathcal{H}_3$.[86]

Since $\mathcal{H}_3*$ is a theorem of the mathematical theory of probability, this reconstruction of Horwich's account is guaranteed to be immune to any formal counterexamples. In this sense, our present reconstruction of Horwich's account is a charitable one. However, even this charitable reconstruction of Horwich's account of CSED has its problems. In the next section, I will briefly discuss some of my remaining worries about Horwich's account of CSED.

3.3.2.5 A Remaining Worry About Horwich's Account

Horwich's $\mathcal{H}_1$ says that a more c-diverse set of evidence E_1 will tend to "rule-out more of the plausible alternative hypotheses $H_{j \neq 1}$" than a less c-diverse set of evidence E_2 will. But, when Horwich gives his canonical curve-fitting examples, he appeals to an *intuitive* sense of diversity (i-diversity) which does not obviously correspond to the formal, *confirmational* diversity specified in $\mathcal{H}_1$. At this point,

[84]This notion of the 'intuitive diversity' (i-diversity) of a data set is never precisely defined by Horwich (1982). But, in canonical curve-fitting contexts, the 'intuitive diversity' of a data set should boil down to some measure of the *spread* (or *variance*) of its abscissa values.

[85]I am being *intentionally* vague here about which kind of diversity Horwich has in mind. I think Horwich has i-diversity in mind; but, he clearly wants this relationship to obtain also with respect to c-diversity. I'll try to resolve this important tension below.

[86]Indeed, as we will show in §A.12, $CP*$ is sufficient for $\mathcal{H}_3$ — *no matter which of our five measures of confirmation is used*! In other words, our charitable reconstruction of Horwich's account of CSED is *completely* insensitive to choice of measure.

as natural question to ask is: "What is the relationship between i-diversity and c-diversity, anyway?"

Ideally, we would like the following *general correspondence* to obtain between the i-diversity and c-diversity of data sets:

($\mathcal{H}_4$) If E_1 is more i-diverse than E_2 in $\mathcal{C}$, then E_1 is more c-diverse than E_2 in $\mathcal{C}$.

If $\mathcal{H}_4$ were generally true (*i.e.*, true for *all* $\mathcal{C}$), then all of the intuitive examples of CSED would automatically translate into formal examples of CSED with just the right mathematical properties. And, Horwich's formal account of CSED (*i.e.*, $\mathcal{H}_1$–$\mathcal{H}_3'$) would be vindicated by its ability to match our intuitions about CSED in all cases. Unfortunately, things don't work out quite this nicely.

It turns out that $\mathcal{H}_4$ is *not* generally true. To see this, let's reconsider Horwich's canonical example of CSED, depicted in Figure 2. In this example, E_1 is more i-diverse *and* more c-diverse than E_2 in $\mathcal{C}$. It is *obvious* why E_1 is more i-diverse than E_2 in $\mathcal{C}$ (just inspect the spread of the abcissa values of E_1 *vs* E_2). However, it is *not* so obvious why E_1 is more c-diverse than E_2 in $\mathcal{C}$. Horwich claims that E_1 tends to rule-out more of the *plausible* alternatives to H_1 than E_2 does. I think it is more perspicuous to say instead that E_1 tends to rule-out more of the *simple* alternatives to H_1 than E_2 does.[87] Horwich doesn't say exactly how we should measure the 'relative simplicity' of competing hypotheses. We can make

[87] Horwich (1982, pages 121-122) and Horwich (1993, pages 66–67) explains that his account of CSED depends on a substantive Bayesian understanding of the *simplicity* of statistical hypotheses. Given our reconstruction of Horwich's account of CSED, we can see vividly why this is so. Horwich seems to be assuming that simple hypotheses have some kind of *a priori probabilistic advantage* over complex hypotheses. This kind of assumption is known as a *simplicity postulate*. Simplicity postulates are a well-known source of controversy in Bayesian philosophy of science. I won't dwell here on the problematical nature of simplicity postulates, since I think they are a problem for a rather large class of Bayesian accounts of CSED. For an interesting discussion of simplicity postulates in Bayesian confirmation theory, see Popper (1980, Appendix **viii*). See, especially, Forster (1995) for a detailed critique of the simplicity postulate in the context of curve-fitting. And, see Kruse (1997) for a refreshing new account of CSED in statistical contexts, which makes explicit and precise the contributions that i-diversity and simplicity play in CSED.

some sense out of Horwich's canonical example, if we make the following plausible and common assumption about how to measure the simplicity of a polynomial hypothesis in a curve-fitting context:

($\mathcal{H}_5$) The simplicity of a polynomial hypothesis H is equal to the dimensionality of the smallest (non-trivial) family of polynomial functions of which H is a member.[88]

If we characterize simplicity in this way, we can *explain why* E_1 tends to rule-out more of the simple alternatives to H_1 than E_2 does in the canonical example depicted in Figure 2. Figure 3 gives us way to picture what's going on in Horwich's canonical example in a rather illuminating and explanatory way.[89]

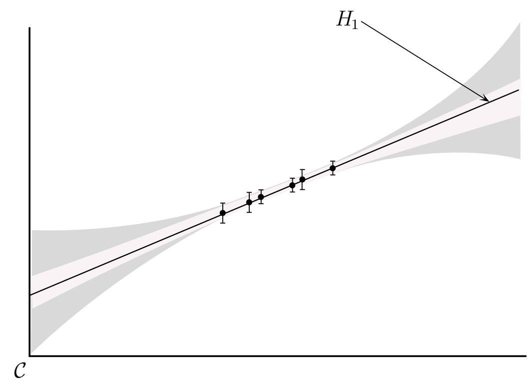

Figure 3: Why Horwich's canonical example has the right formal properties

The lightly shaded area in Figure 3 corresponds to the set of linear hypotheses that are consistent with the data; and, the darkly shaded region corresponds to

[88]This is a standard way of measuring the simplicity of hypotheses in curve-fitting contexts. Take, for instance, the curve H: $y = x^2 + 2x$. The smallest (non-trivial) family of polynomials containing H is family PAR: $y = ax^2 + bx + c$ (where a, b, and c are *adjustable parameters*). The dimensionality of PAR is 3 (which is also the number of freely adjustable parameters in PAR). Hence, the 'simplicity value' of H is 3. As a rule, then, lower dimensionality families contain simpler curves. See Forster and Sober (1994) for more on this notion of simplicity and its important role in modern statistical theory.

[89]Thanks to Malcolm Forster for generating this informative graphic (using *MATHEMATICA*®) and allowing me to use it for this purpose.

the set of parabolic hypotheses that are consistent with the data. Now, if we were to 'spread-out' the abscissa values of the data set in Figure 3 — while keeping the likelihood with respect to H_1 constant, in accordance with $CP*$ — the resulting, more *intuitively* diverse, data set would end up ruling-out more of the plausible (*i.e.*, simpler) alternative hypotheses than the original data set does. This is because the shaded region (whose area is roughly proportional to the number of simple alternatives to that are consistent with the data) will *shrink* as we spread out the data set along the linear H_1. So, in such an example, it is plausible to expect that the more intuitively diverse data set will also be more *confirmationally* diverse in the formal sense of $\mathcal{H}_1$. However, this will *not* generally be the case. In general, *whether or not* $\mathcal{H}_4$ *holds will depend on how complex the hypothesis under test is*. We can imagine situations in which the hypothesis under test is sufficiently complex relative to its competitors in $\mathcal{C}$. In such situations, increasing the spread (or *i*-diversity) of a data set (in accordance with $CP*$) may *not* automatically increase its confirmational diversity.[90]

To see this, consider a confirmational context $\mathcal{C}'$ in which the hypothesis under test H_1 is a highly complex curve, and has only one competitor in $\mathcal{C}'$: a linear hypothesis H_2. Now, assume that some data set E_2 falls exactly on H_1 in such a way that is inconsistent with H_2. If we spread out E_2 in just the right way — in accordance with $CP*$ — to form a more *intuitively* diverse data set E_1, we may end up with a data set that is *not* more confirmationally diverse than E_2. In fact, depending on how complex H_1 is (and how cleverly we choose to spread out E_2 along H_1), E_1 may turn-out to be *less* *c*-diverse than E_2. For instance, E_1 might just happen to fall *exactly* on the linear alternative H_2. This kind of 'non-canonical' confirmational context — in which a more *i*-diverse data set turns out to be *less* *c*-diverse — is pictured below in Figure 4.

To sum up: Horwich's formal sense of confirmational diversity only corresponds

[90]Thanks to Patrick Maher for getting me to see this point clearly.

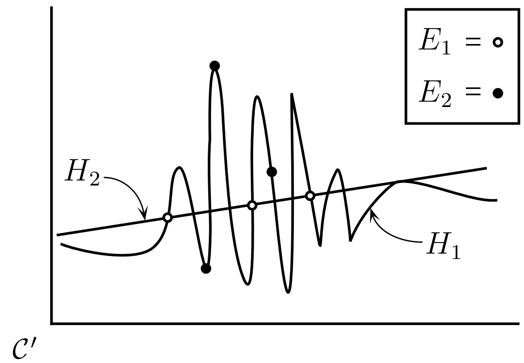

Figure 4: Why the truth of $\mathcal{H}_4$ depends on the complexity of H_1

to his intuitive sense of diversity in contexts where the hypothesis under test is a relatively simple hypothesis. If the hypothesis under test is sufficiently complex relative to its competitors, then the connection between Horwich's formal definition of diversity (in $\mathcal{H}_1$) and the intuitive notion of diversity seen in Horwich's canonical curve-fitting contexts breaks down. Because this connection is essential to the general success of Horwich's approach to explicate our pre-theoretic intuitions about CSED, Horwich's account would seem — at best — to provide an incomplete explication of CSED.

3.3.2.6 The Robustness of Our Reconstruction $\mathcal{H}_3*$

Like Howson and Urbach, Horwich presupposes that the quotient measure r is an adequate Bayesian measure of degree of confirmation. As we have already seen, this is an unfortunate choice of measure. The good news is that, unlike Howson and Urbach's account of CSED, our reconstruction of Horwich's account of CSED (based on $\mathcal{H}_3*$) is *not* sensitive to his choice of measure.[91] In this sense, Horwich's

[91] I reported in §2.2.4 above (see, also, Appendix §A.5) that Horwich's account of CSED trades essentially on (5), and so goes through for d, r, and l (but *not* for s or $\mathfrak{r}$). There is textual evidence that this is an *accurate* reading of Horwich [see Horwich (1982, p. 119)]. But, as I show in §A.12 below, our $\mathcal{H}_3*$-based reconstruction of Horwich's account is *completely robust*. In this sense, our reconstruction based on $\mathcal{H}_3*$ is certainly a more *charitable* one.

account is far more robust than Howson and Urbach's (which, as we show in §A.9, only goes through for measures r and $\mathfrak{r}$), and, perhaps[92] even more robust than our own account, which is based on $\mathcal{D}$. Table 6 summarizes the results reported in this chapter concerning independent evidence and CSED.

	Is $\mathcal{C}$ satisfied by the measure:				
Name and Section of Condition $\mathcal{C}$	d?	r?	l?	s?	$\mathfrak{r}$?
Peirceian Additivity Condition $\mathcal{A}$ (See §3.2.1 and Appendix §A.6)	YES	YES	YES	NO	NO
Negation Symmetry Condition $\mathcal{S}$ (See §3.2.2 and Appendix §A.7)	YES	NO	YES	YES	YES
Conclusive Evidence Condition $\mathcal{K}$ (See §3.2.3)	NO	NO	YES	NO	NO
High-Level Diversity Condition $\mathcal{D}$ (See §3.3, §A.6, and §4.1)	YES	YES	YES	?[93]	?
Low-Level Diversity Condition $\mathcal{D}'$ (See §3.3 (fn. 67) and §4.1)	YES?	YES?	YES	NO	NO
The Urn Condition (UC) (See §3.2.4.2 and Appendix §A.8)	NO	NO	YES	NO	NO
Howson and Urbach's Condition $\mathcal{H}$ (See §3.3.1 and Appendix §A.9)	NO	YES	NO	NO	YES
Horwich's Condition $\mathcal{H}_3*$ (See §3.3.2.4 and Appendix §A.12)	YES	YES	YES	YES	YES

Table 6: Summary of results concerning independent evidence and CSED.

[92]Our account *may* be as robust as our charitable reconstruction of Horwich's account [based on $\mathcal{H}_3*$ rather than (5)]. This question remains open. See §4.1 for more on this open question.

[93]These question-marked answers are either unknown or conjectural. The questions are still open. Computer searches seem to indicate that the "YES?" conjectures are "probable," but no rigorous proofs have yet been found for any of the "?"s. See §4.1 for discussion.

Chapter 4

Future Directions

The direction in which education starts a
man will determine his future life.
— Plato

4.1 Some Remaining Open Questions

The following two important questions remain open:

- Is our high-level diversity desideratum $\mathcal{D}$ satisfied by measures s or $\mathfrak{r}$?
- Which measures satisfy our low-level diversity desideratum $\mathcal{D}'$ (fn. 67)?

($\mathcal{D}'$) If E_1, E_2 individually confirm H, and H screens-off E_1 from E_2, then $\mathfrak{c}(H, E_1 \,\&\, E_2) > \mathfrak{c}(H, E_1)$ and $\mathfrak{c}(H, E_1 \,\&\, E_2) > \mathfrak{c}(H, E_2)$.

We know that d, r, and l satisfy $\mathcal{D}$ (corollary of Theorem 6), and that l satisfies D' (corollary of Theorems 6 and 8). Moreover, we have recently discovered that s and $\mathfrak{r}$ do *not* satisfy D' (countermodels omitted). Computer searches indicate that measures d and r "probably" both satisfy $\mathcal{D}'$, but I have not been able to prove either of these conjectures rigorously. Moreover, I have made little progress on obtaining evidence or proof concerning whether measures s or $\mathfrak{r}$ satisfy $\mathcal{D}$.

These questions are important because they effect the robustness of our account of CSED. If $\mathcal{D}$ is satisfied by all of our measures (*i.e.*, by s and $\mathfrak{r}$), then our account of CSED is more robust than either Howson and Urbach's account ($\mathcal{H}$) or Horwich's (5)-based account, and at least as robust as our charitable reconstruction of Horwich's account (based on $\mathcal{H}_3*$). And, if $\mathcal{D}'$ is satisfied by d and r (as we suspect), then the relation of screening-off is even more central and important in the context of CSED than we have indicated.[94]

4.2 An Analogous Philosophical Problem

Many non-equivalent measures of "distance" or "divergence" between probability distributions have been proposed and defended in the probabilistic and statistical literature.[95] Using different measures can lead to different ordinal judgments about which distributions are "closer" to (or "farther away" from) which. This plurality of divergence measures is particularly important, philosophically, in contexts where such divergences are interpreted as *verisimilitudes*.[96] For example, in the statistical literature on model selection, various criteria have been proposed whose derivations presuppose certain measures of divergence between approximating probability distributions (inferred from data) and true probability distributions.[97] Unfortunately, very few arguments have been presented which aim to narrow this plethora of divergence measures.[98] It would be nice to see a thorough survey of the ramifications of using different measures of divergence in various contexts, and an attempt to provide reasons for favoring one (or few) of these proposals ...

[94]Moreover, if d satisfies $\mathcal{D}'$, then the defenders of d would be able to say that their measure *is* appropriately sensitive to screening-off — to *some* extent. While d is *not* additive in the case of screening-off, at least d says that "two pieces of independent evidence (in the screening-off sense) are better than one." So, defenders of d should be especially interested in this open question.

[95]For an excellent critical survey of this plethora of divergence measures, see Csiszár (1978). These measures are sometimes also called (mutual) *information* measures. As such, they make contact with a very wide variety of problems in mathematics and science [Guiaşu (1977)].

[96]For an encyclopedic survey of the vast landscape of proposed quantitative measures of verisimilitude (both probabilistic and non-probabilistic), see Niiniluoto (1987).

[97]Linhart and Zucchini (1986) show that the choice of divergence measure has a significant effect on the process of deriving model selection criteria. If different divergence measures are used, then different criteria tend to result. Forster and Sober (1994) describe the model selection problem in a broader, philosophical context. Following Akaike (1973), they assume that the Kullback-Leibler divergence measure [Kullback and Leibler (1951)] should be used. Unfortunately, they seem to provide no argument for this foundational assumption about how we should measure the distance between true probability distributions and approximate, epistemic distributions.

[98]The only explicit arguments I have seen (both of which are in favor of the Kullback-Leibler divergence) are presented in Kullback (1997) and Csiszár (1978). And, neither of these arguments seems terribly compelling (form a metaphysical point of view) to me.

Appendix A

Technical Details

The trouble about arguments is, they ain't nothing but
theories, after all, and theories don't prove nothing.
— Mark Twain

A.1 Proof of Theorem 1

Theorem 1. *There exist probability models such that*

$$l(H, E \mid K) \neq l(H \vee E, E \mid K) + l(H \vee \bar{E}, E \mid K).$$

Proof. For simplicity, and without loss of generality (w.l.o.g.) I will assume that $K = \top$ (and, hence, that K can be suppressed from the notation entirely). Then, by the definition of l, we have the following:

$$\begin{aligned} l(H \vee E, E) + l(H \vee \bar{E}, E) &= \log\left[\frac{\Pr(E \mid H \vee E)}{\Pr(E \mid \overline{H \vee E})}\right] + \log\left[\frac{\Pr(E \mid H \vee \bar{E})}{\Pr(E \mid \overline{H \vee \bar{E}})}\right] \\ &= \log\left[\frac{\Pr(E \mid H \vee E)}{\Pr(E \mid \bar{H} \,\&\, \bar{E})}\right] + \log\left[\frac{\Pr(E \mid H \vee \bar{E})}{\Pr(E \mid \bar{H} \,\&\, E)}\right] \\ &= \log\left[\frac{\Pr(E \mid H \vee E)}{0}\right] + \log\left[\frac{\Pr(E \mid H \vee \bar{E})}{1}\right] \\ &= +\infty \\ &\neq l(H, E), \text{ provided only that } l(H, E) \text{ is } \textit{finite}. \end{aligned}$$

There are lots of probability models in which $l(H, E)$ is finite (*i.e.*, models in which $\Pr(E \mid \bar{H}) > 0$). Any one of these is sufficient to establish the desired result.[99] ❑

[99] Readers who balk at the zero denominators in this proof might prefer to carry-out the proof using the (ordinally equivalent) measure l^* of Kemeny and Oppenheim (1952), defined on page 42. Doing so leads to no loss of generality, and eliminates having to fuss with infinities.

A.2 Proof of Theorem 2

Theorem 2. *There exist probability models in which all three of the following obtain: (i)* $H \vDash E$*, (ii)* $r(H \,\&\, X, E \mid K) \neq \Pr(X \mid H \,\&\, K) \cdot r(H, E \mid K)$*, and (iii)* $l(H \,\&\, X, E \mid K) \neq \Pr(X \mid H \,\&\, K) \cdot l(H, E \mid K)$.[100]

Proof. Let K include the information that we are talking about a standard deck of cards with the usual probability structure. Let E be the proposition that some card $\mathcal{C}$, drawn at random from the deck, is a black card (*i.e.*, that $\mathcal{C}$ is either a ♣ or a ♠). Let H be the hypothesis that $\mathcal{C}$ is a ♠. And, let X be the proposition that $\mathcal{C}$ is a 7. Then, we have the following salient probabilities:

$\Pr(X \mid H \,\&\, K) = \frac{1}{13}$	$\Pr(H \mid E \,\&\, K) = \frac{1}{2}$	$\Pr(H \mid K) = \frac{1}{4}$
$\Pr(E \mid H \,\&\, X \,\&\, K) = 1$	$\Pr(E \mid H \,\&\, K) = 1$	$\Pr(E \mid \bar{H} \,\&\, K) = \frac{1}{3}$
$\Pr(H \,\&\, X \mid K) = \frac{1}{52}$	$\Pr(H \,\&\, X \mid E \,\&\, K) = \frac{1}{26}$	$\Pr(E \mid \overline{H \,\&\, X} \,\&\, K) = \frac{25}{51}$

Hence, this probability model is such that all three of the following obtain:

(*i*)
$$H \vDash E$$

(*ii*)
$$\begin{aligned} r(H \,\&\, X, E \mid K) &= \log\left[\frac{1/26}{1/52}\right] \\ &= \log(2) \\ &\neq \Pr(X \mid H \,\&\, K) \cdot r(H, E \mid K) = \frac{1}{13} \cdot \log(2) \end{aligned}$$

[100] Strictly speaking, this theorem is *logically stronger* than Theorem 2, which only requires that there be a probability model in which (*i*) and (*ii*) obtain, and a probability model in which (*i*) and (*iii*) obtain (but, not necessarily *the same* model). Note, also, that the X in my countermodel is, intuitively, an *irrelevant* conjunct. I think this is apropos.

$$(iii) \quad \begin{aligned} l(H \mathbin{\&} X, E \mid K) &= \log \left[\frac{1}{25/51} \right] \\ &= \log \left[\frac{51}{25} \right] \\ &\neq \Pr(X \mid H \mathbin{\&} K) \cdot l(H, E \mid K) = \frac{1}{13} \cdot \log(3) \end{aligned}$$

Consequently, this probability model is sufficient to establish Theorem 2. ❑

A.3 Proof of Theorem 3

Theorem 3. *If E confirms H, and X is confirmationally irrelevant to H, E, and $H \mathbin{\&} E$ (relative to background K), then $\mathfrak{c}(H, E \mid K) > \mathfrak{c}(H \mathbin{\&} X, E \mid K)$, where $\mathfrak{c}$ may be any of our five relevance measures, except r.*

Proof. For the $\mathfrak{c} = d$ case of the theorem, we (again) assume (w.l.o.g.) that $K = \top$, and we reason as follows:

$$\begin{aligned} d(H \mathbin{\&} X, E) &= \Pr(H \mathbin{\&} X \mid E) - \Pr(H \mathbin{\&} X) \quad [\text{def. of } d] \\ &= \frac{\Pr((H \mathbin{\&} E) \mathbin{\&} X)}{\Pr(E)} - \Pr(H \mathbin{\&} X) \quad [\text{def. of } \Pr(\cdot \mid \cdot)] \\ &= \frac{\Pr(H \mathbin{\&} E) \cdot \Pr(X)}{\Pr(E)} - \Pr(H) \cdot \Pr(X) \quad [\text{irrelevance of } X] \\ &= \Pr(X) \cdot [\Pr(H \mid E) - \Pr(H)] \quad [\text{def. of } \Pr(\cdot \mid \cdot), \text{ algebra}] \\ &= \Pr(X) \cdot d(H, E) \quad [\text{def. of } d] \\ &< d(H, E) \quad [\Pr(X) < 1,\ d(H, E) > 0] \end{aligned}$$

The $\mathfrak{c} = s$ and $\mathfrak{c} = \mathfrak{r}$ cases of the theorem follow easily from the above proof for d and the definitions of s and $\mathfrak{r}$, so I will omit the easy proofs in these two cases. That brings us to the $\mathfrak{c} = r$ case of the theorem. For this case, we will prove that $r(H, E) = r(H \mathbin{\&} X, E)$, when X is an irrelevant conjunct. This will also come in handy, below, when we prove the l case of the theorem.

$$r(H \,\&\, X, E) = \frac{\Pr(H \,\&\, X \mid E)}{\Pr(H \,\&\, X)} \qquad \text{[def. of } r\text{]}$$

$$= \frac{\Pr((H \,\&\, E) \,\&\, X)}{\Pr(E) \cdot \Pr(H \,\&\, X)} \qquad \text{[def. of } \Pr(\cdot \mid \cdot)\text{]}$$

$$= \frac{\Pr(H \,\&\, E) \cdot \Pr(X)}{\Pr(E) \cdot \Pr(H) \cdot \Pr(X)} \qquad \text{[irrelevance of } X\text{]}$$

$$= \frac{\Pr(H \mid E)}{\Pr(H)} \qquad \text{[def. of } \Pr(\cdot \mid \cdot)\text{, algebra]}$$

$$= r(H, E) \qquad \text{[def. of } r\text{]}$$

We are now ready to prove the l case of the theorem (*via reductio*).

$$l(H \,\&\, X, E) \geq l(H, E) \qquad \text{[\textit{reductio} assumption]}$$

$$\frac{\Pr(E/H \,\&\, X)}{\Pr(E/\overline{H \,\&\, X})} \geq \frac{\Pr(E/H)}{\Pr(E/\overline{H})} \qquad \text{[def. of } l\text{]}$$

$$\frac{\Pr(H \,\&\, X/E)}{\Pr(H \,\&\, X)} \cdot \frac{\Pr(\overline{H \,\&\, X})}{\Pr(\overline{H \,\&\, X}/E)} \geq \frac{\Pr(H/E)}{\Pr(H)} \cdot \frac{\Pr(\overline{H})}{\Pr(\overline{H}/E)} \qquad \text{[Bayes' Theorem, algebra]}$$

$$r(H \,\&\, X, E) \cdot \frac{\Pr(\overline{H \,\&\, X})}{\Pr(\overline{H \,\&\, X}/E)} \geq r(H, E) \cdot \frac{\Pr(\overline{H})}{\Pr(\overline{H}/E)} \qquad \text{[def. of } r\text{]}$$

$$\frac{\Pr(\overline{H \,\&\, X})}{\Pr(\overline{H \,\&\, X}/E)} \geq \frac{\Pr(\overline{H})}{\Pr(\overline{H}/E)} \qquad [r(H \,\&\, X, E) = r(H, E)\text{, algebra]}$$

$$(1 - \Pr(H/E)) \cdot (1 - \Pr(H \,\&\, X)) \geq (1 - \Pr(H)) \cdot (1 - \Pr(H \,\&\, X/E)) \qquad [\Pr(\overline{X}/Y) = 1 - \Pr(X/Y)\text{, algebra]}$$

$$[d(H \,\&\, X, E) - d(H, E)] + \Pr(H/E) \cdot \Pr(H \,\&\, X) \geq \Pr(H) \cdot \Pr(H \,\&\, X/E) \qquad \text{[def. of } d\text{, algebra]}$$

$$\Pr(H/E) \cdot \Pr(H \,\&\, X) > \Pr(H) \cdot \Pr(H \,\&\, X/E) \qquad [d(H \,\&\, X, E) < d(H, E)\text{, algebra]}$$

$$r(H, E) > r(H \,\&\, X, E) \qquad \text{[def. of } r\text{, algebra]}$$

$$l(H \,\&\, X, E) < l(H, E) \qquad [r(H \,\&\, X, E) = r(H, E)\text{, \textit{reductio}]}$$

This completes the proof of Theorem 3. ❑

A.4 Proof of Theorem 4

Theorem 4. *There exist probability models in which (i)* $\beta \cdot \frac{\Pr(\bar{E} \mid K)}{\Pr(E \mid K)} > \delta$*, (ii)* $\Pr(E \mid K) < \frac{1}{2}$*, and (iii)* $l(H_1, E \mid K) \leq l(H_2, E \mid K)$*. And, there exist probability models in which (i), (ii), and (iv)* $r(H_1, E \mid K) \leq r(H_2, E \mid K)$*.*[101]

Proof. For the l case of the theorem, I will describe a class of probability spaces in which all *four* of the following obtain.[102]

$$(*) \qquad E \text{ confirms both } H_1 \text{ and } H_2 \text{ (given } K)$$

$$(i) \qquad \beta \cdot \frac{\Pr(\bar{E} \mid K)}{\Pr(E \mid K)} > \delta$$

$$(ii) \qquad \Pr(E \mid K) < \frac{1}{2}$$

$$(iii) \qquad l(H_1, E \mid K) < l(H_2, E \mid K)$$

To this end, consider the class of probability spaces containing the three events E, H_1, and H_2 (again, we take $K = \top$, for simplicity and w.l.o.g.) such that the eight basic (or, atomic) events in the space have the following probabilities:

$\Pr(H_1 \,\&\, \bar{H}_2 \,\&\, \bar{E}) = \mathsf{a} = \frac{1169}{17068}$	$\Pr(H_1 \,\&\, H_2 \,\&\, \bar{E}) = \mathsf{d} = \frac{1169}{17068}$
$\Pr(\bar{H}_1 \,\&\, H_2 \,\&\, \bar{E}) = \mathsf{g} = \frac{22913}{85340}$	$\Pr(H_1 \,\&\, \bar{H}_2 \,\&\, E) = \mathsf{b} = \frac{3}{251}$
$\Pr(H_1 \,\&\, H_2 \,\&\, E) = \mathsf{e} = \frac{1}{17}$	$\Pr(\bar{H}_1 \,\&\, H_2 \,\&\, E) = \mathsf{f} = \frac{431}{15060}$
$\Pr(\bar{H}_1 \,\&\, \bar{H}_2 \,\&\, E) = \mathsf{c} = \frac{31}{51204}$	$\Pr(\bar{H}_1 \,\&\, \bar{H}_2 \,\&\, \bar{E}) = \mathsf{h} = \frac{42203}{85340}$

Now, we verify that the class of probability spaces described above is such that $(*)$, (i), (ii), and (iii) all obtain. To see that (iii) holds, note that we have

[101] Where $\beta =_{df} \Pr(H_1 \,\&\, E \mid K) - \Pr(H_2 \,\&\, E \mid K)$, and $\delta =_{df} \Pr(H_1 \,\&\, \bar{E} \mid K) - \Pr(H_2 \,\&\, \bar{E} \mid K)$.

[102] It crucial that our countermodel be such that $(*)$ obtains. For instance, if we were to allow E to confirm H_2 but *dis*confirm H_1, then "counterexamples" would be easy to find, but they would not be a problem for Eells's resolution of the Grue Paradox, since Eells is clearly talking about cases in which E (the observation of a large number of green emeralds, before t_0) confirms *both* H_1 (that all emeralds are green) *and* H_2 (that all emeralds are grue).

$l(H_1, E) < l(H_2, E)$, by the following computation.

$$l(H_1, E) = \log\left[\frac{(1 - \mathsf{a} - \mathsf{b} - \mathsf{d} - \mathsf{e})\ (\mathsf{b} + \mathsf{e})}{(\mathsf{a} + \mathsf{b} + \mathsf{d} + \mathsf{e})\ (\mathsf{c} + \mathsf{f})}\right] = \log\left(\frac{20418220}{2210931}\right) \approx \log(9.24)$$
$$l(H_2, E) = \log\left[\frac{(1 - \mathsf{d} - \mathsf{e} - \mathsf{f} - \mathsf{g})\ (\mathsf{e} + \mathsf{f})}{(\mathsf{d} + \mathsf{e} + \mathsf{f} + \mathsf{g})\ (\mathsf{b} + \mathsf{c})}\right] = \log\left(\frac{3298925933}{349345115}\right) \approx \log(9.44)$$

To see that (ii) holds, note that $\Pr(E) < \frac{1}{2}$.

$$\Pr(E) = \mathsf{b} + \mathsf{c} + \mathsf{e} + \mathsf{f} = \frac{1}{10} = 0.1$$

To see that (i) holds, note that $\beta \cdot \frac{\Pr(\bar{E})}{\Pr(E)} > \delta$.

$$\beta \cdot \frac{\Pr(\bar{E})}{\Pr(E)} = (\mathsf{b} - \mathsf{f}) \cdot \frac{9/10}{1/10} = -\frac{1}{60} \cdot 9 = -\frac{3}{20} = -0.15$$
$$\delta = \mathsf{a} - \mathsf{g} = -\frac{1}{5} = -0.2$$

Finally, $(*)$ holds in our example, since $l(H_1, E) > 0$ and $l(H_2, E) > 0$.

For the r case of the theorem, I will describe a class of probability spaces in which all *four* of the following obtain.

$(*)$ $\quad E$ confirms both H_1 and H_2 (given K)

(i) $$\beta \cdot \frac{\Pr(\bar{E} \mid K)}{\Pr(E \mid K)} > \delta$$

(ii) $$\Pr(E \mid K) < \frac{1}{2}$$

(iii) $$r(H_1, E \mid K) = r(H_2, E \mid K)$$

To this end, consider the class of probability spaces containing the three events E, H_1, and H_2 (again assuming $K = \top$, for simplicity and w.l.o.g.) such that the eight basic (or, atomic) events in the space have the following probabilities:

$\Pr(H_1 \,\&\, \bar{H}_2 \,\&\, \bar{E}) = \mathsf{a} = \frac{167}{12253020}$	$\Pr(H_1 \,\&\, H_2 \,\&\, \bar{E}) = \mathsf{d} = \frac{1}{2049}$
$\Pr(\bar{H}_1 \,\&\, H_2 \,\&\, \bar{E}) = \mathsf{g} = \frac{636509}{91652589600}$	$\Pr(H_1 \,\&\, \bar{H}_2 \,\&\, E) = \mathsf{b} = \frac{5}{299}$
$\Pr(H_1 \,\&\, H_2 \,\&\, E) = \mathsf{e} = \frac{1}{15}$	$\Pr(\bar{H}_1 \,\&\, H_2 \,\&\, E) = \mathsf{f} = \frac{4201}{269100}$
$\Pr(\bar{H}_1 \,\&\, \bar{H}_2 \,\&\, E) = \mathsf{c} = \frac{269}{269100}$	$\Pr(\bar{H}_1 \,\&\, \bar{H}_2 \,\&\, \bar{E}) = \mathsf{h} = \frac{82440714571}{91652589600}$

Now, we verify that the class of probability spaces described above is such that $(*)$, (i), (ii), and (iii) all obtain. To see that (iii) holds, note that we have $r(H_1, E) = r(H_2, E)$, by the following computation.

$$r(H_1, E) = \log\left[\frac{\mathsf{b}+\mathsf{e}}{(\mathsf{a}+\mathsf{b}+\mathsf{d}+\mathsf{e})\;(\mathsf{b}+\mathsf{c}+\mathsf{e}+\mathsf{f})}\right] = \log\left(\frac{2992}{301}\right)$$
$$r(H_2, E) = \log\left[\frac{\mathsf{e}+\mathsf{f}}{(\mathsf{b}+\mathsf{c}+\mathsf{e}+\mathsf{f})\;(\mathsf{d}+\mathsf{e}+\mathsf{f}+\mathsf{g})}\right] = \log\left(\frac{2292}{301}\right)$$

To see that (ii) holds, note that $\Pr(E) < \frac{1}{2}$.

$$\Pr(E) = \mathsf{b}+\mathsf{c}+\mathsf{e}+\mathsf{f} = \frac{1}{10} = 0.1$$

To see that (i) holds, note that $\beta \cdot \frac{\Pr(\bar{E})}{\Pr(E)} > \delta$.

$$\beta \cdot \frac{\Pr(\bar{E})}{\Pr(E)} = (\mathsf{b}-\mathsf{f}) \cdot \frac{9/10}{1/10} = \frac{1}{900} \cdot 9 = \frac{1}{100} = 0.1$$
$$\delta = \mathsf{a}-\mathsf{g} = -\frac{1}{149600} \approx 6.68 \times 10^{-6}$$

Finally, $(*)$ holds in our example, since $r(H_1, E) > 0$ and $r(H_2, E) > 0$.

This completes the proof of Theorem 4. ❑

A.5 Proof of Theorem 5

Theorem 5. *There exist probability models in which all three of the following obtain: (i)* $\Pr(H \mid E_1 \,\&\, K) > \Pr(H \mid E_2 \,\&\, K)$*, (ii)* $\mathfrak{r}(H, E_1 \mid K) \leq \mathfrak{r}(H, E_2 \mid K)$*, and* *(iii)* $s(H, E_1 \mid K) \leq s(H, E_2 \mid K)$.

Proof. I will prove Theorem 5 by describing a class of probability spaces in which all *four* of the following obtain.[103]

$$(*) \qquad \text{Each of } E_1 \text{ and } E_2 \text{ confirms } H \text{ (given } K)$$

$$(i) \qquad \Pr(H \mid E_1 \,\&\, K) > \Pr(H \mid E_2 \,\&\, K)$$

$$(ii) \qquad \mathfrak{r}(H, E_1 \mid K) < \mathfrak{r}(H, E_2 \mid K)$$

$$(iii) \qquad s(H, E_1 \mid K) < s(H, E_2 \mid K)$$

To this end, consider the class of probability spaces containing the three events E_1, E_2, and H (again letting $K = \top$, for simplicity and w.l.o.g.) such that the eight basic (or, atomic) events in the space have the following probabilities:

$\Pr(E_1 \,\&\, \bar{E}_2 \,\&\, \bar{H}) = \mathsf{a} = \frac{1}{1000}$	$\Pr(E_1 \,\&\, E_2 \,\&\, \bar{H}) = \mathsf{b} = \frac{1}{1000}$
$\Pr(\bar{E}_1 \,\&\, E_2 \,\&\, \bar{H}) = \mathsf{c} = \frac{1}{200}$	$\Pr(E_1 \,\&\, \bar{E}_2 \,\&\, H) = \mathsf{d} = \frac{1}{100}$
$\Pr(E_1 \,\&\, E_2 \,\&\, H) = \mathsf{e} = \frac{1}{100}$	$\Pr(\bar{E}_1 \,\&\, E_2 \,\&\, H) = \mathsf{f} = \frac{1}{25}$
$\Pr(\bar{E}_1 \,\&\, \bar{E}_2 \,\&\, H) = \mathsf{g} = \frac{1}{500}$	$\Pr(\bar{E}_1 \,\&\, \bar{E}_2 \,\&\, \bar{H}) = \mathsf{h} = \frac{931}{1000}$

Now, we verify that the class of probability spaces described above is such that $(*)$, (i), (ii), and (iii) all obtain. To see that $(*)$ and (i) both hold, note that we

[103] It is important that our countermodel satisfy $(*)$. In the ravens paradox, it should be granted that both a black raven (E_1) and a non-black non-raven (E_2) may confirm that all ravens are black (H). Similarly, it should be granted that both a "varied" (or "diverse") set of evidence (E_1) and a "narrow" set of evidence (E_2) can confirm a hypothesis under test (H). Wayne (1995) presents a "counterexample" to Horwich's (1982) Bayesian account of evidential diversity which fails to respect this constraint. See Fitelson (1996) and §3.3.2.2 above for details.

have $\Pr(H \mid E_1) > \Pr(H)$, $\Pr(H \mid E_2) > \Pr(H)$, and $\Pr(H \mid E_1) > \Pr(H \mid E_2)$:

$$\Pr(H \mid E_1) = \frac{\mathsf{d}+\mathsf{e}}{\mathsf{a}+\mathsf{b}+\mathsf{d}+\mathsf{e}} = \frac{10}{11} \approx 0.909$$
$$\Pr(H \mid E_2) = \frac{\mathsf{e}+\mathsf{f}}{\mathsf{b}+\mathsf{c}+\mathsf{e}+\mathsf{f}} = \frac{25}{28} \approx 0.893$$
$$\Pr(H) = \mathsf{d}+\mathsf{e}+\mathsf{f}+\mathsf{g} = \frac{31}{500} = 0.062$$

To see that (*ii*) holds, note that $\mathfrak{r}(H, E_1) < \mathfrak{r}(H, E_2)$.

$$\mathfrak{r}(H, E_1) = (\mathsf{a}+\mathsf{b}+\mathsf{d}+\mathsf{e}) \cdot \left[\frac{\mathsf{d}+\mathsf{e}}{\mathsf{a}+\mathsf{b}+\mathsf{d}+\mathsf{e}} - (\mathsf{d}+\mathsf{e}+\mathsf{f}+\mathsf{g})\right] = \frac{4659}{250000} \approx 0.0186$$
$$\mathfrak{r}(H, E_2) = (\mathsf{b}+\mathsf{c}+\mathsf{e}+\mathsf{f}) \cdot \left[\frac{\mathsf{e}+\mathsf{f}}{\mathsf{b}+\mathsf{c}+\mathsf{e}+\mathsf{f}} - (\mathsf{d}+\mathsf{e}+\mathsf{f}+\mathsf{g})\right] = \frac{727}{15625} \approx 0.0465$$

Finally, to see that (*iii*) holds, note that $s(H, E_1) < s(H, E_2)$.[104]

$$s(H, E_1) = \frac{\mathsf{d}+\mathsf{e}}{\mathsf{a}+\mathsf{b}+\mathsf{d}+\mathsf{e}} - \frac{\mathsf{f}+\mathsf{g}}{1-\mathsf{a}-\mathsf{b}-\mathsf{d}-\mathsf{e}} = \frac{1553}{1793} \approx 0.866$$
$$s(H, E_2) = \frac{\mathsf{e}+\mathsf{f}}{\mathsf{b}+\mathsf{c}+\mathsf{e}+\mathsf{f}} - \frac{\mathsf{d}+\mathsf{g}}{1-\mathsf{b}-\mathsf{c}-\mathsf{e}-\mathsf{f}} = \frac{727}{826} \approx 0.880$$

This completes the proof of Theorem 5. ❑

A.6 Proof of Theorem 6

Theorem 6. *Each of the measures d, r, and l satisfies $\mathcal{A}$, but s and $\mathfrak{r}$ violate $\mathcal{A}$.*

[104]This is also a model in which both $\Pr(E_1 \mid H) - \Pr(E_1) < \Pr(E_2 \mid H) - \Pr(E_2)$, and $\Pr(E_1 \mid H) - \Pr(E_1 \mid \bar{H}) < \Pr(E_2 \mid H) - \Pr(E_2 \mid \bar{H})$ (check this!). So, the relevance measures of both Mortimer (1988, §11.1) and Nozick (1981, 252), respectively, *also* violate (5).

Proof. This proof has five parts.[105] The proofs for d and r are easy:

$$
\begin{aligned}
d(H, E_1 \mid E_2) &= d(H, E_1) \\
\therefore\ \Pr(H \mid E_1 \,\&\, E_2) - \Pr(H \mid E_2) &= \Pr(H \mid E_1) - \Pr(H) \\
\therefore\ \Pr(H \mid E_1 \,\&\, E_2) - \Pr(H) &= (\Pr(H \mid E_1) - \Pr(H)) \\
&\quad + (\Pr(H \mid E_2) - \Pr(H)) \\
\therefore\ d(H, E_1 \,\&\, E_2) &= d(H, E_1) + d(H, E_2)
\end{aligned} \tag{d}
$$

$$
\begin{aligned}
r(H, E_1 \mid E_2) &= r(H, E_1) \\
\therefore\ \log[\Pr(H \mid E_1 \,\&\, E_2)] - \log[\Pr(H \mid E_2)] &= \log[\Pr(H \mid E_1)] - \log[\Pr(H)] \\
\therefore\ \log[\Pr(H \mid E_1 \,\&\, E_2)] - \log[\Pr(H)] &= (\log[\Pr(H \mid E_1)] - \log[\Pr(H)]) \\
&\quad + (\log[\Pr(H \mid E_2)] - \log[\Pr(H)]) \\
\therefore\ r(H, E_1 \,\&\, E_2) &= r(H, E_1) + r(H, E_2)
\end{aligned} \tag{r}
$$

The proof for l is only slightly more involved. For the l case of the theorem, we will prove that the likelihood ratio (λ) is *multiplicative* under the assumption of confirmational independence. That the *logarithm* of λ (*i.e.*, l) is *additive* under

[105]Throughout this part of the Appendix, we will suppress the contents of the background evidence K other than E_1 and E_2. Moreover, we will try to prove the strongest results we know. Usually, these will be considerably stronger than the theorems that are stated in the main text.

the assumption of confirmational independence then follows straightaway.

$$
(l) \quad
\begin{aligned}
& l(H, E_1 \mid E_2) = l(H, E_1) \\
\therefore\ & \lambda(H, E_1 \mid E_2) = \lambda(H, E_1) \qquad \text{[strict monotonicity of } \log(\bullet)\text{]} \\
\therefore\ & \frac{\Pr(E_1 \mid H \,\&\, E_2)}{\Pr(E_1 \mid \bar{H} \,\&\, E_2)} = \frac{\Pr(E_1 \mid H)}{\Pr(E_1 \mid \bar{H})} \qquad \text{[def. of } \lambda\text{]} \\
\therefore\ & \frac{\Pr(E_1 \mid H)}{\Pr(E_1 \mid \bar{H})} = \frac{\Pr(E_1 \,\&\, E_2 \mid H)}{\Pr(E_1 \,\&\, E_2 \mid \bar{H})} \cdot \frac{\Pr(E_2 \mid \bar{H})}{\Pr(E_2 \mid H)} \qquad \text{[def. of } \Pr(\bullet \mid \bullet)\text{]} \\
\therefore\ & \frac{\Pr(E_1 \,\&\, E_2 \mid H)}{\Pr(E_1 \,\&\, E_2 \mid \bar{H})} = \frac{\Pr(E_1 \mid H)}{\Pr(E_1 \mid \bar{H})} \cdot \frac{\Pr(E_2 \mid H)}{\Pr(E_2 \mid \bar{H})} \\
\therefore\ & \lambda(H, E_1 \,\&\, E_2) = \lambda(H, E_1) \cdot \lambda(H, E_2) \\
\therefore\ & l(H, E_1 \,\&\, E_2) = l(H, E_1) + l(H, E_2) \qquad \text{[additivity of } \log(\bullet)\text{]}
\end{aligned}
$$

The s case of the theorem is far trickier, because it requires us to show that there is *no* (symmetric) isotone function f such that, for all E_1, E_2, and H, if E_1 and E_2 are confirmationally independent regarding H according to s, then $s(H, E_1 \,\&\, E_2) = f[s(H, E_1), s(H, E_2)]$, where f is linear in some (isotonically) transformed space. Happily, I have proven the following *much stronger* result:

($*$) There exist probability models $\mathcal{M}_1$ and $\mathcal{M}_2$ such that:

$\mathcal{M}_1$	$\mathcal{M}_2$
$s(H, E_1 \mid E_2) = s(H, E_1) = \frac{1}{4}$	$s(H, E_1 \mid E_2) = s(H, E_1) = \frac{1}{4}$
$s(H, E_2 \mid E_1) = s(H, E_2) = \frac{1}{4}$	$s(H, E_2 \mid E_1) = s(H, E_2) = \frac{1}{4}$
$s(H, E_1 \,\&\, E_2) = \frac{15}{44} - \frac{96}{4451 + 3 \cdot \sqrt{1254641}}$ ≈ 0.3286	$s(H, E_1 \,\&\, E_2) = \frac{15}{44} + \frac{96}{3 \cdot \sqrt{1254641} - 4451}$ ≈ 0.2529

Of course, it follows from ($*$) that there can be *no function f* ***whatsoever*** such that for all E_1, E_2, and H, if E_1 and E_2 are confirmationally independent regarding H according to s, then $s(H, E_1 \,\&\, E_2) = f[s(H, E_1), s(H, E_2)]$. This is because ($i$) $\mathcal{M}_1$ and $\mathcal{M}_2$ are both such that E_1 and E_2 are confirmationally independent regarding H according to s, (ii) In $\mathcal{M}_1$ and $\mathcal{M}_2$, $s(H, E_1)$ and $s(H, E_2)$ are *constant* at the same value of $\frac{1}{4}$, but (iii) The value of $s(H, E_1 \,\&\, E_2)$ in $\mathcal{M}_1$ is different from the value of $s(H, E_1 \,\&\, E_2)$ in $\mathcal{M}_2$. So, whatever $s(H, E_1 \,\&\, E_2)$ is in cases where E_1 and E_2 are confirmationally independent regarding H according to s, it cannot (in general) be of the form $f[s(H, E_1), s(H, E_2)]$ *for any f whatsoever*, since functions cannot give different values for identical arguments. I will not display here all the calculations necessary to show that the models $\mathcal{M}_1$ and $\mathcal{M}_2$ reported below have the desired properties.[106]

$\mathcal{M}_1$	
$\Pr(H \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{1}{100}$	$\Pr(H \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{1}{1000}$
$\Pr(\bar{H} \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{1}{100}$	$\Pr(H \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{1}{1000}$
$\Pr(H \,\&\, E_1 \,\&\, E_2) = \frac{9\cdot(1183-\sqrt{1254641})}{70400}$	$\Pr(\bar{H} \,\&\, E_1 \,\&\, E_2) = \frac{87\cdot(1183-\sqrt{1254641})}{352000}$
$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{1}{100}$	$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{121}{125} + \frac{3\cdot(\sqrt{1254641}-1183)}{8000}$

$\mathcal{M}_2$	
$\Pr(H \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{1}{100}$	$\Pr(H \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{1}{1000}$
$\Pr(\bar{H} \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{1}{100}$	$\Pr(H \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{1}{1000}$
$\Pr(H \,\&\, E_1 \,\&\, E_2) = \frac{9\cdot(1183+\sqrt{1254641})}{70400}$	$\Pr(\bar{H} \,\&\, E_1 \,\&\, E_2) = \frac{87\cdot(1183+\sqrt{1254641})}{352000}$
$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{1}{100}$	$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{121}{125} - \frac{3\cdot(\sqrt{1254641}+1183)}{8000}$

The $\mathfrak{r}$ case can be established in a similar manner, by showing that there are probability models $\mathcal{M}_3$ and $\mathcal{M}_4$ such that:

[106] The probability models in this Appendix were discovered using MATHEMATICA® algorithms written by the author. See section A.14 below for more on these techniques.

$\mathcal{M}_3$	$\mathcal{M}_4$
$\mathfrak{r}(H, E_1 \mid E_2) = \mathfrak{r}(H, E_1) = \frac{1}{4}$	$\mathfrak{r}(H, E_1 \mid E_2) = \mathfrak{r}(H, E_1) = \frac{1}{4}$
$\mathfrak{r}(H, E_2 \mid E_1) = \mathfrak{r}(H, E_2) = \frac{1}{4}$	$\mathfrak{r}(H, E_2 \mid E_1) = \mathfrak{r}(H, E_2) = \frac{1}{4}$
$\mathfrak{r}(H, E_1 \,\&\, E_2) = \frac{627}{3128}$	$\mathfrak{r}(H, E_1 \,\&\, E_2) = \frac{3121}{15570}$

I will not display here all the calculations necessary to show that the models $\mathcal{M}_3$ and $\mathcal{M}_4$ reported below have the desired properties.

$\mathcal{M}_3$	
$\Pr(H \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{1}{768}$	$\Pr(H \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{1}{192}$
$\Pr(\bar{H} \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{70026913}{200066880}$	$\Pr(H \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{1}{192}$
$\Pr(H \,\&\, E_1 \,\&\, E_2) = \frac{2}{7}$	$\Pr(\bar{H} \,\&\, E_1 \,\&\, E_2) = \frac{1402}{1458821}$
$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{70026913}{200066880}$	$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{1257121}{800267520}$

$\mathcal{M}_4$	
$\Pr(H \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{1}{768}$	$\Pr(H \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{1}{256}$
$\Pr(\bar{H} \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{368581709}{1052947200}$	$\Pr(H \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{1}{256}$
$\Pr(H \,\&\, E_1 \,\&\, E_2) = \frac{2}{7}$	$\Pr(\bar{H} \,\&\, E_1 \,\&\, E_2) = \frac{100378}{28791525}$
$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{368581709}{1052947200}$	$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{557861}{350982400}$

This completes the proof of Theorem 6. ❑

A.7 Proof of Theorem 7

Theorem 7. *Each of the measures d, l, s, and $\mathfrak{r}$ satisfies $\mathcal{S}$, but r violates $\mathcal{S}$.*

Proof. This proof has five parts. The d, l, s, and $\mathfrak{r}$ cases reduce to trivial algebraic identities (I won't include here the easy proofs for these cases). For the

r case, we need to show that there exists a probability model $\mathcal{M}$ such that: (i) both $r(H, E_1 \mid E_2) = r(H, E_1)$ and $r(H, E_2 \mid E_1) = r(H, E_2)$, but ($ii$) either $r(\bar{H}, E_1 \mid E_2) \neq r(\bar{H}, E_1)$ or $r(\bar{H}, E_2 \mid E_1) \neq r(\bar{H}, E_2)$. Here is one such model $\mathcal{M}$ (computational details omitted).

$\mathcal{M}$	
$\Pr(H \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{1}{64}$	$\Pr(H \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{1}{64}$
$\Pr(\bar{H} \,\&\, E_1 \,\&\, \bar{E}_2) = \frac{1}{64}$	$\Pr(H \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{87+\sqrt{66265}}{704}$
$\Pr(H \,\&\, E_1 \,\&\, E_2) = \frac{1}{4}$	$\Pr(\bar{H} \,\&\, E_1 \,\&\, E_2) = \frac{1}{16}$
$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, E_2) = \frac{1}{8}$	$\Pr(\bar{H} \,\&\, \bar{E}_1 \,\&\, \bar{E}_2) = \frac{276-\sqrt{66265}}{704}$

This completes the proof of Theorem 7. ❑

A.8 Proof of Theorem 8

Theorem 8. *The measures d, r, s, and $\mathfrak{r}$ violate* (UC)*, but l satisfies* (UC)*.*

Proof. For the d, r, s, and $\mathfrak{r}$ cases of the theorem, it will suffice to produce an urn example (*i.e.*, an assignment of values on $(0, 1)$ to the variables x, y, and z) such that either $\mathfrak{c}(H, W_1 \mid W_2) \neq \mathfrak{c}(H, W_1)$ or $\mathfrak{c}(H, W_2 \mid W_1) \neq \mathfrak{c}(H, W_2)$, for each of the four measures d, r, s, and $\mathfrak{r}$. The following (far from extreme[107]) assignment does the trick: $\langle x, y, z\rangle = \langle \frac{1}{2}, \frac{49}{100}, \frac{1}{2}\rangle$. On this assignment, we have the following salient

[107] y can be *arbitrarily close to* $\frac{1}{2}$, while preserving the counterexample. See footnote 61.

probabilistic facts (computational details omitted):

$$(d) \qquad d(H, W_2 \mid W_1) = 2450/485199 < d(H, W_2) = 1/198$$

$$(r) \qquad r(H, W_2 \mid W_1) = \log(4950/4901) < r(H, W_2) = \log(100/99)$$

$$(s) \qquad s(H, W_2 \mid W_1) = 245000/24500099 < s(H, W_2) = 100/9999$$

$$(\mathfrak{r}) \qquad \mathfrak{r}(H, W_2 \mid W_1) = 49/80000 < \mathfrak{r}(H, W_2) = 1/400$$

For the l case, we will show that l satisfies the stronger condition (SC).

$$(l) \qquad \begin{aligned} &\Pr(E_1 \mid H \,\&\, E_2) = \Pr(E_1 \mid H) && \text{[screening-off assumption]} \\ &\Pr(E_1 \mid \bar{H} \,\&\, E_2) = \Pr(E_1 \mid \bar{H}) && \text{[screening-off assumption]} \\ \therefore\ &\frac{\Pr(E_1 \mid H \,\&\, E_2)}{\Pr(E_1 \mid \bar{H} \,\&\, E_2)} = \frac{\Pr(E_1 \mid H)}{\Pr(E_1 \mid \bar{H})} \\ &\therefore\ l(H, E_1 \mid E_2) = l(H, E_1) \end{aligned}$$

It is easy to show that, for any of the three measures d, r, or l (but *not* for s or $\mathfrak{r}$), $\mathfrak{c}(H, E_1 \mid E_2) = \mathfrak{c}(H, E_1)$ *iff* $\mathfrak{c}(H, E_2 \mid E_1) = \mathfrak{c}(H, E_2)$. That, together with the reasoning above, completes the l case, and with it the proof of Theorem 8. ❑

Interestingly, l does *not* satisfy the *converse* of (SC) [or (UC)]. The following model is one in which both: (i) E_1 and E_2 are confirmationally independent regarding H according to l (*i.e.*, $l(H, E_i \mid E_j) = l(H, E_i)$ $[i \neq j]$), but (ii) H does *not* screen-off E_1 from E_2 (*i.e.*, $\Pr(E_1 \,\&\, E_2 \mid H) \neq \Pr(E_1 \mid H) \cdot \Pr(E_2 \mid H)$). This explains why I choose *not* to assume screening-off as a *necessary* condition for confirmational independence (computational details omitted).

$\mathcal{M}$	
$\Pr(H \;\&\; \bar{E}_1 \;\&\; \bar{E}_2) = \frac{683}{3800}$	$\Pr(H \;\&\; E_1 \;\&\; \bar{E}_2) = \frac{1}{50}$
$\Pr(\bar{H} \;\&\; E_1 \;\&\; \bar{E}_2) = \frac{1169}{1102000}$	$\Pr(H \;\&\; \bar{E}_1 \;\&\; E_2) = \frac{1}{50}$
$\Pr(H \;\&\; E_1 \;\&\; E_2) = \frac{1}{456}$	$\Pr(\bar{H} \;\&\; E_1 \;\&\; E_2) = \frac{2}{29}$
$\Pr(\bar{H} \;\&\; \bar{E}_1 \;\&\; E_2) = \frac{922}{1305}$	$\Pr(\bar{H} \;\&\; \bar{E}_1 \;\&\; \bar{E}_2) = \frac{15179}{9918000}$

A.9 Proof of Theorem 9

Theorem 9. *$\mathcal{H}$ is true if $\mathfrak{c} = r$ or $\mathfrak{c} = \mathfrak{r}$, but $\mathcal{H}$ is false if $\mathfrak{c} = d$, $\mathfrak{c} = l$, or $\mathfrak{c} = s$.*

Proof. For the r case of the theorem, we begin by assuming that the probabilistic '*ceteris paribus* clause' (CP) is satisfied. That is, we assume: $\Pr(E_1 \mid H) = \Pr(E_2 \mid H) = \Pr(E_1 \;\&\; E_2 \mid H) = 1$. Then, we apply (CP), the definition of r, and Bayes' Theorem to derive the following pair of probabilistic facts:

$$
\begin{aligned}
r(H, E_2 \mid E_1) &= \log\left[\frac{\Pr(H \mid E_1 \;\&\; E_2)}{\Pr(H \mid E_1)}\right] \\
&= \log\left[\frac{\Pr(E_1 \;\&\; E_2 \mid H)\cdot\Pr(H)\cdot\Pr(E_1)}{\Pr(E_1 \;\&\; E_2)\cdot\Pr(E_1 \mid H)\cdot\Pr(H)}\right] \\
&= \log\left[\frac{\Pr(E_1)}{\Pr(E_1 \;\&\; E_2)}\right] \\
&= \log\left[\frac{1}{\Pr(E_2 \mid E_1)}\right]
\end{aligned}
\tag{23}
$$

$$
\begin{aligned}
r(H, E_2) &= \log\left[\frac{\Pr(H \mid E_2)}{\Pr(H)}\right] \\
&= \log\left[\frac{\Pr(E_2 \mid H)\cdot\Pr(H)}{\Pr(E_2)\cdot\Pr(H)}\right] \\
&= \log\left[\frac{1}{\Pr(E_2)}\right]
\end{aligned}
\tag{24}
$$

Finally, we assume that E_1 and E_2 are positively correlated under Pr. Or, more formally, we assume that $\Pr(E_2 \mid E_1) > \Pr(E_2)$. In conjunction with facts (23) and (24) above, this yields $r(H, E_2 \mid E_1) < r(H, E_2)$, as desired.[108] ❑

For the $\mathfrak{r}$ case, we reason as follows:

$$\begin{aligned}\mathfrak{r}(H, E_2 \mid E_1) &= \Pr(H \,\&\, E_1 \,\&\, E_2) \cdot \Pr(E_1) - \Pr(H \,\&\, E_1) \cdot \Pr(E_1 \,\&\, E_2) \quad [\text{def. } \mathfrak{r}] \\ &= \Pr(H) \cdot \Pr(E_1) - \Pr(H) \cdot \Pr(E_1 \,\&\, E_2) \quad [(CP)] \\ &= \Pr(H) \cdot [\Pr(E_1) - \Pr(E_1 \,\&\, E_2)] \quad [\text{algebra}]\end{aligned}$$

and

$$\begin{aligned}\mathfrak{r}(H, E_2) &= \Pr(H \,\&\, E_2) - \Pr(H) \cdot \Pr(E_2) \quad [\text{def. } \mathfrak{r}] \\ &= \Pr(H) - \Pr(H) \cdot \Pr(E_2) \quad [(CP)] \\ &= \Pr(H) \cdot [1 - \Pr(E_2)] \quad [\text{algebra}]\end{aligned}$$

Some algebra, and another application of (CP), yield:

$$\text{Given } (CP),\ \mathfrak{r}(H, E_2 \mid E_1) < \mathfrak{r}(H, E_2) \text{ iff } \Pr(H) > 0.$$

But, we can safely assume that $\Pr(H) > 0$ in the cases of interest. So, to be charitable to Howson and Urbach, we must conclude that (in all interesting cases in which (CP) holds), their condition $\mathcal{H}$ is satisfied by Carnap's measure $\mathfrak{r}$. ❑

For the d, l, and s cases, it will suffice to produce a probability model in which (*i*) $\Pr(E_1 \mid H) = \Pr(E_2 \mid H) = \Pr(E_1 \,\&\, E_2 \mid H) = 1$, (*ii*) $\Pr(E_2 \mid E_1) > \Pr(E_2)$, but (*iii*) $\mathfrak{c}(H, E_2 \mid E_1) \geq \mathfrak{c}(H, E_2)$, for $\mathfrak{c} = d$, $\mathfrak{c} = l$, and $\mathfrak{c} = s$. The following example

[108] Notice that Howson and Urbach's claim that (CP) can be weakened even further to (CP') $\Pr(E_1 \mid H) = \Pr(E_2 \mid H)$ — while still preserving the truth of the $\mathfrak{c} = r$ case of Theorem 9 — is *false*. If we only assume (CP'), then we will need to establish that $\Pr(E_2 \mid E_1) > \frac{\Pr(E_2)}{\Pr(E_2 \mid H)}$, in order to prove that $r(H, E_2 \mid E_1) < r(H, E_2)$. Unfortunately, $\Pr(E_2 \mid E_1) > \frac{\Pr(E_2)}{\Pr(E_2 \mid H)}$ does *not* follow from the fact that E_1 and E_2 are positively correlated under Pr, unless one also assumes that $\Pr(E_2 \mid H) = 1$, which brings us (essentially) back to (CP). Explicit countermodels can be produced (omitted). A similar result can be shown for Carnap's measure $\mathfrak{r}$ (proof omitted).

does the trick. A card is drawn at random from a standard deck. Let H be the hypothesis that the card is the Q♠, E_1 be the proposition that the card is either a 10 or a face card, and E_2 be the proposition that the card is either a ♡, or the Q♠, or the 9♠. I omit the calculations which show that this example has the desired properties (*i*)–(*iii*) listed above. This completes the proof of Theorem 9. ❑

A.10 Proofs for Milne's Desiderata (7)–(11)

Milne's (7) is just our $\mathcal{R}$, which is *trivially* satisfied by all five of our measures d, r, l, s, and $\mathfrak{r}$. Milne's (8) is almost as trivial, since all five of our measures are defined using simple arithmetic functions of the left and right sides of the inequalities listed way back on page 5. This makes it is easy to show that each of our five measures d, r, l, s, and $\mathfrak{r}$ is "some function of the values $\Pr(\cdot \mid K)$ and $\Pr(\cdot \mid \cdot \& K)$ assumed on the at most sixteen truth-functional combinations of E and H," as (8) requires. Since (7) and (8) are so obvious and uncontroversial, I won't bother to prove them.

It *is* worth proving that Milne's (9) is satisfied by each of our five measures d, r, l, s, and $\mathfrak{r}$. This is not so obvious, and it will also prove very useful in §A.12 below, when we prove the measure insensitivity of our charitable reconstruction ($\mathcal{H}_3*$) of Horwich's account of CSED. I will not prove both clauses of (9), but only the second clause, which states that:

(9) If $\Pr(E \mid H \,\&\, K) = \Pr(E' \mid H \,\&\, K)$ and $\Pr(E \mid K) < \Pr(E' \mid K)$ then $\mathfrak{c}(H, E \mid K) \geq \mathfrak{c}(H, E' \mid K)$.

The first clause of (9) can be proved in a similar way. I choose to prove the second clause only, because it is the second clause of (9) that will be the crucial lemma in our proof of the measure insensitivity of ($\mathcal{H}_3*$) in §A.12, below. We begin by proving the d case of (9), as follows. For simplicity, and without loss of generality, we take $K = \top$, and so suppress K from the notation.

$$\begin{aligned}
\Pr(E \mid H) &= \Pr(E' \mid H) && \text{[assumption]}\\
\frac{\Pr(H \mid E) \cdot \Pr(E)}{\Pr(H)} &= \frac{\Pr(H \mid E') \cdot \Pr(E')}{\Pr(H)} && \text{[Bayes' Theorem]}\\
\Pr(H \mid E) \cdot \Pr(E) &= \Pr(H \mid E') \cdot \Pr(E') && \text{[algebra]}\\
\Pr(H \mid E) &> \Pr(H \mid E') && [\Pr(E) < \Pr(E'), \text{ algebra}]\\
d(H, E) &> d(H, E') && [\text{def. } d, \text{ algebra}] \quad \square
\end{aligned}$$

Next, we prove the s case of (9). For this proof, we will make use of the fact (proved during the d case, above), that the assumptions of (9) entail $\Pr(H \mid E) > \Pr(H \mid E')$. Given this lemma, all we need to show for the s case is that the assumptions of (9) also entail $\Pr(H \mid \bar{E}) < \Pr(H \mid \bar{E}')$. From this (together with the definition of s and simple algebra), it will follow that $s(H, E) > s(H, E')$.[109] So, we will now prove that the assumptions of (9) entail $\Pr(H \mid \bar{E}) < \Pr(H \mid \bar{E}')$. During this proof, I will make use of the following (easy) theorem from the probability calculus: $(\mathcal{T})$ $\Pr(\bar{X} \mid Y) = 1 - \Pr(X \mid Y)$.

$$\begin{aligned}
\Pr(E \mid H) &= \Pr(E' \mid H) && \text{[assumption]}\\
\Pr(\bar{E} \mid H) &= \Pr(\bar{E}' \mid H) && [\mathcal{T}, \text{ algebra}]\\
\Pr(\bar{E} \mid H) \cdot \Pr(H) &= \Pr(\bar{E}' \mid H) \cdot \Pr(H) && \text{[algebra]}\\
\frac{\Pr(\bar{E} \mid H) \cdot \Pr(H)}{\Pr(\bar{E})} &< \frac{\Pr(\bar{E}' \mid H) \cdot \Pr(H)}{\Pr(\bar{E}')} && [\Pr(E) < \Pr(E'), \mathcal{T}, \text{ algebra}]\\
\Pr(H \mid \bar{E}) &< \Pr(H \mid \bar{E}') && \text{[Bayes' Theorem]} \quad \square
\end{aligned}$$

[109] Recall from algebra that, for all $x, y, z, u \in [0, 1]$, if $x > z$ and $y < u$, then $x - y > z - u$. Letting $x = \Pr(H \mid E)$, $y = \Pr(H \mid \bar{E})$, $z = \Pr(H \mid E')$, and $u = \Pr(H \mid \bar{E}')$, yields the desired result that if $\Pr(H \mid E) > \Pr(H \mid E')$ and $\Pr(H \mid \bar{E}) < \Pr(H \mid \bar{E}')$, then $s(H, E) > s(H, E')$.

Now, we're ready for the $\mathfrak{r}$ case of (9). First, recall that (with $K = \top$):

$$\mathfrak{r}(H, E) = \Pr(H \,\&\, E) - \Pr(H) \cdot \Pr(E)$$

So, we must prove that the assumptions of (9) entail:

$$\Pr(H \,\&\, E) - \Pr(H) \cdot \Pr(E) > \Pr(H \,\&\, E') - \Pr(H) \cdot \Pr(E')$$

By simple algebra, this inequality holds iff

$$\Pr(H \,\&\, E) - \Pr(H \,\&\, E') > \Pr(H) \cdot [\Pr(E) - \Pr(E')]$$

Dividing both sides by $\Pr(H)$ and applying the definition of conditional probability yields the following equivalent inequality:

$$\Pr(E \mid H) - \Pr(E' \mid H) > \Pr(E) - \Pr(E')$$

But, since $\Pr(E \mid H) = \Pr(E' \mid H)$ is an assumption of (9), all we need to prove now is that the assumptions of (9) entail:

$$0 > \Pr(E) - \Pr(E')$$

But, by simple algebra, this inequality holds iff

$$\Pr(E) < \Pr(E')$$

which is just the other assumption of (9). ❑

The l case of (9) is all that remains. For this case, we must prove that the assumptions of (9) entail:

$$\frac{\Pr(E \mid H)}{\Pr(E \mid \bar{H})} > \frac{\Pr(E' \mid H)}{\Pr(E' \mid \bar{H})}$$

Since $\Pr(E \mid H) = \Pr(E' \mid H)$ is an assumption of (9), this means all we have to show is that the assumptions of (9) entail:

$$\frac{1}{\Pr(E \mid \bar{H})} > \frac{1}{\Pr(E' \mid \bar{H})}$$

By simple algebra, this is equivalent to showing that the assumptions of (9) entail:

$$\Pr(E \mid \bar{H}) < \Pr(E' \mid \bar{H})$$

This is an easy consequence of the two assumptions of (9). To see this, note that, by the law of total probability:

$$\Pr(E) = \Pr(E \mid H) \cdot \Pr(H) + \Pr(E \mid \bar{H}) \cdot \Pr(\bar{H})$$
$$\Pr(E') = \Pr(E' \mid H) \cdot \Pr(H) + \Pr(E' \mid \bar{H}) \cdot \Pr(\bar{H})$$

The inequality $\Pr(E \mid \bar{H}) < \Pr(E' \mid \bar{H})$ then follows straightaway from these two facts, together with the two assumptions of (9). This completes the proof that all five of our measures d, r, l, s, and $\mathfrak{r}$ satisfy Milne's (9). The robustness of (9) will be used again later to establish the robustness of ($\mathcal{H}_3*$) in §A.12, below. ❑

Next, we need to show that the measures s and $\mathfrak{r}$ do *not* satisfy the (second clause of) Milne's desideratum (10).[110] For this, it will be sufficient (but not

[110]It is easy to show that d, r, and l *do* satisfy (10) (proofs omitted).

necessary) to produce a probability model in which all four of the following obtain:

(i) $$s(H, E_1 \mathbin{\&} E_2) = 0$$

(ii) $$\mathfrak{r}(H, E_1 \mathbin{\&} E_2) = 0$$

(iii) $$s(H, E_1) + s(H, E_2 \mid E_1) \neq 0$$

(iv) $$\mathfrak{r}(H, E_1) + \mathfrak{r}(H, E_2 \mid E_1) \neq 0$$

The following probability model does the trick (computations omitted). ❑

$\Pr(H \mathbin{\&} \bar{E}_1 \mathbin{\&} \bar{E}_2) = \frac{59617}{239263440}$	$\Pr(H \mathbin{\&} E_1 \mathbin{\&} \bar{E}_2) = \frac{1}{480}$
$\Pr(\bar{H} \mathbin{\&} E_1 \mathbin{\&} \bar{E}_2) = \frac{1}{32}$	$\Pr(H \mathbin{\&} \bar{E}_1 \mathbin{\&} E_2) = \frac{1}{32}$
$\Pr(H \mathbin{\&} E_1 \mathbin{\&} E_2) = \frac{1}{5552}$	$\Pr(\bar{H} \mathbin{\&} E_1 \mathbin{\&} E_2) = \frac{1}{194}$
$\Pr(\bar{H} \mathbin{\&} \bar{E}_1 \mathbin{\&} E_2) = \frac{1}{2}$	$\Pr(\bar{H} \mathbin{\&} \bar{E}_1 \mathbin{\&} \bar{E}_2) = \frac{3833159}{8917792}$

Finally, we need to show that the measures d, l, s, and $\mathfrak{r}$ do *not* satisfy Milne's desideratum (11).[111] For this purpose, it will be sufficient (but not necessary) to produce a probability model in which all five of the following obtain:

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

(ii) $$d(H, E) \neq d(H', E)$$

(iii) $$l(H, E) \neq l(H', E)$$

(iv) $$s(H, E) \neq s(H', E)$$

(v) $$\mathfrak{r}(H, E) \neq \mathfrak{r}(H', E)$$

The following probability model does the trick (computations omitted). ❑

[111] It is easy to show that r *does* satisfy (11). See Milne (1996).

$\Pr(H \,\&\, \bar{H}' \,\&\, \bar{E}) = \frac{1}{8}$	$\Pr(H \,\&\, H' \,\&\, \bar{E}) = \frac{1}{16}$
$\Pr(\bar{H} \,\&\, H' \,\&\, \bar{E}) = \frac{1}{16}$	$\Pr(H \,\&\, \bar{H}' \,\&\, E) = \frac{11}{48}$
$\Pr(H \,\&\, H' \,\&\, E) = \frac{1}{12}$	$\Pr(\bar{H} \,\&\, H' \,\&\, E) = \frac{1}{8}$
$\Pr(\bar{H} \,\&\, \bar{H}' \,\&\, E) = \frac{1}{6}$	$\Pr(\bar{H} \,\&\, \bar{H}' \,\&\, \bar{E}) = \frac{7}{48}$

This completes the proofs concerning Milne's desiderata (9)–(11). ❑

A.11 Proofs of Wayne's (20), (21), and (22)

A.11.1 Proof of (20)

The task at hand is to prove:

(20) In $\mathcal{C}_w$, E_1 is more c-diverse than E_2.

Proof. Recall that, in Wayne's counterexample context $\mathcal{C}_w$, the hypothesis under test H_1 has only two competitors with non-negligible priors: H_2 and H_3. Moreover, Wayne stipulates that, in $\mathcal{C}_w$, both:

$$\begin{aligned}\Pr(E_1 \mid H_2) &= 0.4\\ &< \Pr(E_2 \mid H_2) = 0.5\text{, and}\\ \Pr(E_1 \mid H_3) &= 0.4\\ &< \Pr(E_2 \mid H_3) = 0.6.\end{aligned}$$

In conjunction with the characterization of c-diversity given in $\mathcal{H}_1$, these two facts about $\mathcal{C}_w$ yield the desired result. ❑

A.11.2 Proof of (21)

We need to demonstrate that:

(21) In $\mathcal{C}_w$, E_2 confirms H_1 more strongly than E_1 confirms H_1, according to the ratio measure r (*i.e.*, $r(H_1, E_2) > r(H_1, E_1)$).

Proof. Wayne's description of $\mathcal{C}_w$, together with Bayes's Theorem, and the definition of r reported in $\mathcal{H}_2$ yields:

$$\begin{aligned} r(H_1, E_1) &= \frac{\Pr(E_1 \mid H_1)}{\sum_i \Pr(E_1 \mid H_i) \cdot \Pr(H_i)} \\ &= \frac{0.2}{(0.2 \cdot 0.2) + (0.4 \cdot 0.2) + (0.4 \cdot 0.6)} \\ &\approx 0.555, \text{ and} \\ r(H_1, E_2) &= \frac{\Pr(E_2 \mid H_1)}{\sum_i \Pr(E_2 \mid H_i) \cdot \Pr(H_i)} \\ &= \frac{0.6}{(0.6 \cdot 0.2) + (0.5 \cdot 0.2) + (0.6 \cdot 0.6)} \\ &\approx 1.034. \end{aligned}$$

Hence, we have $r(H_1, E_2) > r(H_1, E_1)$ in $\mathcal{C}_w$, as desired. ❑

A.11.3 Proof of (22)

Next, we will prove:

(22) In $\mathcal{C}_w$, E_2 confirms H_1; whereas, E_1 *dis*confirms H_1.

Proof. According to Bayesian confirmation theory, $E \,\substack{\text{confirms} \\ \textit{dis}\text{confirms}}\, H$ if and only if $r(H, E) \gtrless 1$. This fact about Bayesian confirmation theory, in conjunction with the following two facts about $\mathcal{C}_w$ (both of which were proved in the preceding section of this Appendix):

$$r(H_1, E_1) \approx 0.555 < 1, \text{ and } r(H_1, E_2) \approx 1.034 > 1,$$

yields the desired result. ❑

A.12 Proof of the Robustness of $\mathcal{H}_3*$

In this section, we will not only prove $\mathcal{H}_3*$, which presupposes the *ratio* measure r of degree of confirmation; we will also show that $\mathcal{H}_3*$ *is true for all five of our measures of confirmation d, r, l, s, and* $\mathfrak{r}$. In other words, our charitable reconstruction of Horwich's account of CSED is *completely insensitive* to choice of measure of confirmation. That is, we have the following robust result:

> If both E_1 is more c-diverse than E_2 in $\mathcal{C}$, and $\Pr(E_1 \mid H_1) = \Pr(E_2 \mid H_1)$ in $\mathcal{C}$, then $\mathfrak{c}(H, E_1) > \mathfrak{c}(H, E_2)$ in $\mathcal{C}$, where $\mathfrak{c}$ may be *any* of our five measures of confirmation d, r, l, s, or $\mathfrak{r}$.

Proof. It turns out that this is an immediate corollary of the robustness of Milne's desideratum (9) (proved in §A.10). We explain why, as follows. From the nature of confirmational contexts, we know that $\bar{H}_1$ is logically equivalent to $\bigvee H_{i\neq 1}$, where the $H_{i\neq 1}$ are mutually exclusive. Hence, from the probability calculus, we have:

$$\Pr(E_1 \mid \bar{H}_1) = \frac{\sum_{i\neq 1} \Pr(E_1 \mid H_i) \cdot \Pr(H_i)}{\sum_{i\neq 1} \Pr(H_i)},$$

and

$$\Pr(E_2 \mid \bar{H}_1) = \frac{\sum_{i\neq 1} \Pr(E_2 \mid H_i) \cdot \Pr(H_i)}{\sum_{i\neq 1} \Pr(H_i)}.$$

From which (with some algebraic manipulation), we may obtain:

$$(**) \qquad (\forall i \neq 1)[\Pr(E_1 \mid H_i) < \Pr(E_2 \mid H_i)] \Longrightarrow \Pr(E_1 \mid \bar{H}_1) < \Pr(E_2 \mid \bar{H}_1)$$

But, the antecedent of $(**)$ just says that E_1 is more c-diverse than E_2. Therefore, if E_1 is more c-diverse than E_2, then $\Pr(E_1 \mid \bar{H}_1) < \Pr(E_2 \mid \bar{H}_1)$. The other assumption of $\mathcal{H}_3*$ is that $\Pr(E_1 \mid H_1) = \Pr(E_2 \mid H_1)$. By the law of total probability, these two assumptions of $\mathcal{H}_3*$ entail $\Pr(E_1) < \Pr(E_2)$. So, the assumptions of $\mathcal{H}_3*$ entail the assumptions of Milne's (9). Therefore, as an immediate corollary of (9), we get $\mathfrak{c}(H, E_1) > \mathfrak{c}(H, E_2)$ — *for all five relevance measures* d*,* r*,* l*,* s*, and* $\mathfrak{r}$*.* ❑

A.13 Counterexample to $CP_1 \Longrightarrow \mathcal{H}_3$

In this section, we show (by generating a concrete counterexample) that:

$$CP_1 \not\Rightarrow \mathcal{H}_3.$$

Proof. Consider a simple context[112] $\mathcal{C}_{w_1}$ in which only three hypotheses have substantial prior probabilities, $\Pr(H_1) = 0.2$, $\Pr(H_2) = 0.2$, $\Pr(H_3) = 0.6$, and two data sets E_1 and E_2 such that:

$$\begin{array}{ll} \Pr(E_1 \,|\, H_1) = 0.41 & \Pr(E_2 \,|\, H_1) = 0.6 \\ \Pr(E_1 \,|\, H_2) = 0.4 & \Pr(E_2 \,|\, H_2) = 0.5 \\ \Pr(E_1 \,|\, H_3) = 0.4 & \Pr(E_2 \,|\, H_3) = 0.6 \end{array}$$

This is plainly a case in which E_1 is more c-diverse than E_2, in the sense of $\mathcal{H}_1$: for all H_i, $\Pr(E_1 \,|\, H_i)$ is significantly less than $\Pr(E_2 \,|\, H_i)$. Moreover, this is also a case in which the probabilistic 'ceteris paribus clause' CP_1 holds. As the following calculations show, both E_1 and E_2 confirm H_1 in $\mathcal{C}_{w_1}$.

$$\begin{aligned} \Pr(H_1 \,|\, E_1) &= \frac{\Pr(E_1 \,|\, H_1) \cdot \Pr(H_1)}{\sum_i \Pr(E_1 \,|\, H_i) \cdot \Pr(H_i)} \\ &= \frac{0.41 \cdot 0.2}{(0.41 \cdot 0.2) + (0.4 \cdot 0.2) + (0.4 \cdot 0.6)} \\ &\approx 0.204 \\ &> \Pr(H_1) = 0.2 \end{aligned}$$

[112] Note: $\mathcal{C}_{w_1}$ is just a slight modification of Wayne's $\mathcal{C}_w$. I have just changed the value of $\Pr(E_1 \,|\, H_1)$ in Wayne's $\mathcal{C}_w$ from 0.2 to 0.41, while leaving the rest of $\mathcal{C}_w$ unchanged.

and

$$\begin{aligned}
\Pr(H_1 \mid E_2) &= \frac{\Pr(E_2 \mid H_1) \cdot \Pr(H_1)}{\sum_i \Pr(E_2 \mid H_i) \cdot \Pr(H_i)} \\
&= \frac{0.6 \cdot 0.2}{(0.6 \cdot 0.2) + (0.5 \cdot 0.2) + (0.6 \cdot 0.6)} \\
&\approx 0.207 \\
&> \Pr(H_1) = 0.2.
\end{aligned}$$

Finally, $\mathcal{C}_{w_1}$ is such that E_2 (the *less* c-diverse collection of evidence) confirms H_1 more strongly than E_1 (the *more* c-diverse collection of evidence), according to all three Bayesian relevance measures r, d, and l. This follows from the fact that $\Pr(H_1 \mid E_2) > \Pr(H_1 \mid E_1)$ in $\mathcal{C}_{w_1}$ (see above), and the proofs given in the previous section concerning the sufficiency of $\Pr(H_1 \mid E_2) > \Pr(H_1 \mid E_1)$ for $\mathfrak{c}(H_1, E_2) > \mathfrak{c}(H_1, E_1)$, where $\mathfrak{c}$ is any of the three Bayesian relevance measures of confirmation r, d, or l. Therefore, $\mathcal{C}_{w_1}$ is a counterexample to $CP_1 \Longrightarrow \mathcal{H}_3$. ❑

A.14 Using *MATHEMATICA*® to Reason About the Probability Calculus[113]

The Kolmogorov axioms for (finite) probability calculus have some very convenient properties. One of these is the fact that any set of statements in the Kolmogorov probability calculus (*i.e.*, any finite set of inequalities involving a Kolmogorov probability function Pr defined over a finite number n of atomic events) can be translated into a logically equivalent set of algebraic inequalities involving only the unconditional probabilities of the 2^n logical combinations of the n events in the space. Moreover, for small probability spaces (*e.g.*, spaces with 3 or fewer atomic events, which includes all probability spaces needed for almost all examples in this monograph), this translation is easy to write down and carry out. Once this translation is carried out, determining whether a set of statements involving Pr is satisfiable is then just a matter of determining whether a set of inequalities on the simplex of the corresponding $2^n - 1$-dimensional Euclidean space has a solution. Happily, version 4 of *MATHEMATICA*® [Wolfram (1999) is *the* reference on *MATHEMATICA*®] has a very powerful built-in inequality solver, which is well suited to exactly these kinds of problems. This allows us to write *MATHEMATICA*® algorithms which will — in a surprisingly wide variety of cases — verify non-trivial theorems and find non-trivial counterexamples in (small) Kolmogorov probability spaces.

I will briefly discuss how this can be done in the case of 3-element Kolmogorov probability spaces.[114] First, it helps to picture a typical 3-event probability space Ω, using a Venn diagram, as in Figure 5 below. The three atomic events are called X, Y, and Z, and the unconditional probabilities of the $2^3 = 8$ logical combinations of these events are denoted by a, b, c, d, e, f, g and h. It is easiest to think of the

[113] The material in this section is taken from Fitelson (2001c).

[114] I have had limited success at finding models and verifying theorems in spaces up to size 4. Various ways of optimizing the computations involved here are discussed in Fitelson (2001c).

a, b, c, d, e, f, g and h as the *areas* of the 8 logically distinct regions in the Venn diagram. A probability space is specified simply by assigning real numbers on $[0, 1]$ to a, b, c, d, e, f, g and h, where the only constraint on these numbers is that they must sum to 1. Now, the translation from statements in the probability calculus involving Pr, X, Y, and Z into algebraic inequalities involving a, b, c, d, e, f, g and h is quite simple. The basic idea is that the unconditional probability of any event α in the space is just the *sum* of whichever of the a, b, c, d, e, f, g, and h are contained in α (*i.e.*, the area of the region corresponding to α in the Venn diagram). Conditional probabilities are defined in the standard (Kolmogorov) way, in terms of the unconditional probabilities. In *MATHEMATICA*®, this is all very easy to encode. What follows (next page) is the output from a *MATHEMATICA*® version 4 session notebook, which explains how this encoding can be carried out and applied to non-trivial examples.[115]

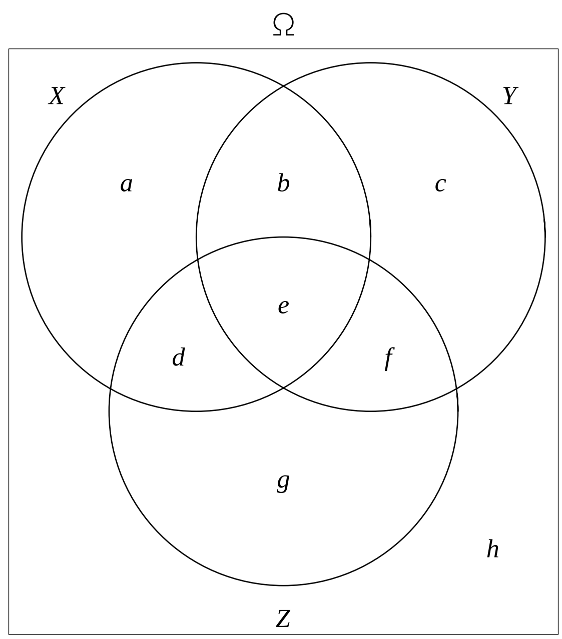

Figure 5: Venn diagram visualization of a 3-event probability space Ω

[115] All relevant *MATHEMATICA*® notebooks and packages pertaining to this monograph are available upon request from the author.

First, we'll need to load in one of *Mathematica*'s standard add-on packages, called **`Notation`**. This package will allow us to use traditional and familiar notation and syntax for our user-defined functions.

```
In[1]:= Needs[Utilities‘Notation‘];
```

With the **`Notation`** package loaded, the following few simple lines of *Mathematica* code suffice to define the unconditional and conditional probability functions **`Pr[•]`** and **`Pr[• | •]`** for a generic 3-event Kolmogorov probability space. Here, (because it is easier to implement these functions using set-theoretic rather than propositional structures) we use Ω (rather than ⊤) to denote the necessarily true proposition, and we use "¬**X**" rather than "$\bar{X}$" to denote logical negation. Also, we use the set-theoretic connectives "∪" and "∩" rather than their propositional counterparts "∨" and "&".

```
In[2]:= Ω = {{a}, {b}, {c}, {d}, {e}, {f}, {g}, {h}};
        X = {{a}, {b}, {d}, {e}};
        Y = {{b}, {c}, {e}, {f}};
        Z = {{d}, {e}, {f}, {g}};
        Notation[¬α_ ⟺ Complement[Ω, α_]];
        Pr[α_] := Plus@@Flatten[α ∪ {0}]//.h → 1 − (a + b + c + d + e + f + g);
In[3]:= Notation[Pr[α_ | β_] ⟺ Pr[α_, β_]]
In[4]:= Pr[α_ | β_] := Pr[α ∩ β] / Pr[β];
```

Here are a couple of examples to illustrate how effortlessly our *Mathematica* code can translate arbitrary statements or expressions in the probability calculus into their corresponding algebraic equivalents.

```
In[5]:=  Pr[(¬X) ∪ Y]
Out[5]= 1 − a − d

In[6]:=  Pr[X | ¬(Y ∩ Z)] == Pr[X | Z]
```

Out[6]= $\frac{a + b + d}{1 - e - f} == \frac{d + e}{d + e + f + g}$

We can easily verify that **Pr[•]** and **Pr[• | •]** satisfy the probability axioms. First, we should have **Pr[Ω]** = **1**, and **Pr[¬Ω]** = **0**. This is easily verified, as follows:

```
In[7]:= Pr[Ω] == 1
Out[7]= True
```

```
In[8]:= Pr[¬Ω] == 0
Out[8]= True
```

The (general) addition law, **Pr[X ∪ Y]** = **Pr[X]**+ **Pr[Y]** – **Pr[X ∩ Y]**, is easily verified as follows:

```
In[9]:= Pr[X∪Y] == Pr[X] + Pr[Y] - Pr[X∩Y]
Out[9]= True
```

Other simple theorems (involving only equalities) can be verified easily. Here's a very easy theorem:

```
In[10]:= Pr[X] + Pr[¬X] == 1
Out[10]= True
```

And, Bayes' Theorem is also very easy to verify:

```
In[11]:= Pr[X|Y] == Pr[Y|X] Pr[X] / Pr[Y]
Out[11]= True
```

The theorem of total probability is also very simple:

```
In[12]:= Pr[X] == Pr[X | Y] Pr[Y] + Pr[X | ¬Y] Pr[¬Y]
Out[12]= True
```

Things get more interesting if we want to prove (or *dis*prove) various *implications* in probability calculus. For instance, what if we want to show that *if* **Pr[X]** = **1**, *then* **Pr[X | Y]** = **Pr[X]**? For this, we will need something a bit more powerful. *Mathematica* (version 4) has a built-in function (currently, in the **Developer** context) called

`InequalityInstance[]` which will, for any set of inequalities (or equations) of the kind we are investigating (*i.e.*, involving polynomials with real coefficients) try to find a satisfying assignment of values. Now, this procedure (based on Tarskian quantifier elimination) will not work for sufficiently complex systems of inequalities. However, for most problems in small probability spaces such as these, the kinds of systems of inequalities that are of interest do tend to be tractable for *Mathematica's* **`InequalityInstance`** function. Of course, if **`InequalityInstance`** finds a model, then we know the system of inequalities is solvable. Hence, in such a case, we know that the corresponding set of statements in the probability calculus is *satisfiable*. And, in most cases (*i.e.*, unless *Mathematica* gives us an error), if no model is found by **`InequalityInstance`**, then this means that the set is *unsatisfiable*. This gives us, more or less, a decision procedure for (sufficiently simple) sets of statements in probability calculus. Let's try our example. We want to show that if **`Pr[X]`** = **`1`**, then **`Pr[X | Y]`** = **`Pr[X]`**. So, let's use **`InequalityInstance`** to look for a probability model in which **`Pr[X]`** = **`1`**, but **`Pr[X | Y]`** ≠ **`Pr[X]`**. First, we need to tell *Mathematica* what makes an assignment to the eight variables **`a`** … **`h`** a *probability* assignment. This is easily represented as the following conjunction of inequalities, which says that **`a`** … **`h`** are all on [0,1], and that the sum of the **`a`** … **`h`** is equal to 1:

```
In[13]:= prob = 0 ≤ a ≤ 1&&0 ≤ b ≤ 1&&0 ≤ c ≤ 1&&0 ≤ d ≤ 1&&0 ≤ e ≤ 1&&
           0 ≤ f ≤ 1&&0 ≤ g ≤ 1&&0 ≤ h ≤ 1&&a + b + c + d + e + f + g + h == 1;

In[14]:= Developer`InequalityInstance[
          prob&&Pr[X] == 1&&Pr[X | Y] ≠ Pr[X], {a, b, c, d, e, f, g, h}]
Out[14]= {}
```

Mathematica quickly tells us that this is impossible. Therefore, **`Pr[X]`** = **`1`** implies **`Pr[X | Y]`** = **`Pr[X]`**. Here's a slightly more involved example. Let's show that **`Pr[X]`** = **`Pr[Y]`** does *not* imply **`Pr[X ∩ Z]`** = **`Pr[Y ∩ Z]`**.

```
In[15]:= Developer`InequalityInstance[
           prob&&Pr[X] == Pr[Y]&&Pr[X ∩ Z] ≠ Pr[Y ∩ Z],
           {a, b, c, d, e, f, g, h}]
Out[15]= {a → 1/2, b → 0, c → 0, d → 0, e → 0, f → 1/2, g → 0, h → 0}
```

That was easy! Here's a much more difficult example. It is well known that the *pairwise* independence of three events does *not* imply their independence *per se*. We can try to prove this in one fell swoop, using:

```
In[16]:= Developer`InequalityInstance[
           prob&&Pr[X ∩ Y] == Pr[X] Pr[Y]&& Pr[Y ∩ Z] == Pr[Y] Pr[Z]&&
             Pr[X ∩ Z] == Pr[X] Pr[Z]&&Pr[(X∩Y) ∩Z] ≠ Pr[X] Pr[Y] Pr[Z],
           {a, b, c, d, e, f, g, h}]
Out[16]= Abort[]
```

But, this will take a *very long* time (and *a lot* of memory) to find a model. Here's a way to make the computation much easier. First, let's use *Mathematica*'s **`Solve`** function to find a (generic) solution (in terms of **`a`** … **`h`**) of the three equations which state pairwise independence. [I have also added in the constraint **`Pr[X]`** = **`1/3`**, to further simplify the search space.]

```
In[17]:= pairwise =
         FullSimplify[Solve[Pr[X∩Y] == Pr[X] Pr[Y]&&
         Pr[Y∩Z] == Pr[Y] Pr[Z]&&Pr[X∩Z] == Pr[X] Pr[Z]&&Pr[X] == 1/3][[1]]]
Out[17]= {c → b (2 - 9 d - 9 e) + 3 e - 9 e (d + e),
           g → d (2 - 9 b - 9 e) + 3 e - 9 e (b + e),
           a → 1/3 - b - d - e, f → -e + 9 (b + e) (d + e)}
```

This allows us to reduce the number of variables in the problem from 7 to 4. Under this assumption, the denial of independence reduces to:

```
In[18]:= ind = Pr[(X∩Y) ∩Z] ≠ Pr[X] Pr[Y] Pr[Z]//.pairwise//FullSimplify
Out[18]= e ≠ 3 (b + e) (d + e)
```

Therefore, all we need is a model in which **prob** and **ind** are satisfied, under the assumption of **pairwise**. The following command allows us to find values of **b**, **d**, **e**, and **h** that solve this problem.

```
In[19]:= m1 =
         Developer`InequalityInstance[(prob&&ind)//.pairwise, {b, d, e, h}]
```

Out[19]= $\left\{e \to \frac{1}{258}, d \to \frac{3}{28}, b \to \frac{1}{133}, h \to \frac{17738141}{41314056}\right\}$

Putting this together with the assignments to the other 3 variables in **pairwise**, yields the following total probability model:

```
In[20]:= model = Join[pairwise//.m1, m1]
```

Out[20]= $\left\{c \to \frac{210429}{13771352}, g \to \frac{2954345}{13771352}, a \to \frac{14741}{68628}, f \to \frac{310241}{41314056}, e \to \frac{1}{258}, d \to \frac{3}{28}, b \to \frac{1}{133}, h \to \frac{17738141}{41314056}\right\}$

We can verify that **model** has all the right properties, as follows. First, it is a *probability* model:

```
In[21]:= prob//.model
Out[21]= True
```

Second, it's a model on which X, Y, and Z are *pairwise* independent.

```
In[22]:= (Pr[X ∩ Y] == Pr[X] Pr[Y]&&
             Pr[Y ∩ Z] == Pr[Y] Pr[Z]&& Pr[X ∩ Z] == Pr[X] Pr[Z])//.model
Out[22]= True
```

And, finally, it's a model on which X, Y, and Z are *not* independent *per se*.

```
In[23]:= Pr[(X∩Y)∩Z] ≠ Pr[X] Pr[Y] Pr[Z]//.model
Out[23]= True
```

By using these sorts of techniques, I was able to find all the models reported in preceding sections of the Appendix, and to verify many of the theorems proved above. As a finalé, I will now demonstrate how we can easily verify one of the key theorems in this monograph using *Mathematica*. First, we use the **Notation** package to define the likelihood ratio measure *l* in its familiar syntactical form:

```
In[24]:= Notation[l[x_, y_ | z_] ⟺ Pr[y_, x_ ∩ z_]/Pr[y_, (¬x_) ∩ z_]]
         Notation[l[x_, y_] ⟺ Pr[y_, x_]/Pr[y_, (¬x_)]]
```

Next we make a slight change of notation, to make the problem look more familiar:

```
In[25]:= ℍ = X; 𝔼1 = Y; 𝔼2 = Z;
```

We will now verify that *l* satisfies the (SC) condition. First, we ask *Mathematica* to **Solve** for the (generic) conditions **SO** under which screening-off (of 𝔼1 from 𝔼2 by ℍ) holds:

```
In[26]:= SO = FullSimplify[Solve[Pr[𝔼1∩𝔼2 | ℍ] == Pr[𝔼1 | ℍ] Pr[𝔼2 | ℍ] &&
                 Pr[𝔼1∩𝔼2 | ¬ℍ] == Pr[𝔼1 | ¬ℍ] Pr[𝔼2 | ¬ℍ]][[4]]]
```

Out[26]= $\left\{c \to -\frac{f\,(d^2 + a\,e + d\,(-1 + a + e + f + g))}{d\,(f + g)},\ b \to \frac{a\,e}{d}\right\}$

Then, we encode the definition of confirmational independence **CI** (for *l)*, as follows:

```
In[27]:= CI = l[ℍ, 𝔼1 | 𝔼2] == l[ℍ, 𝔼1] ;
```

Finally, we ask *Mathematica* to find a probability model in which **SO** is true but **CI** is false:

```
In[28]:= Developer‘InequalityInstance[
           (prob&&!CI)//.SO, {a, b, c, d, e, f, g, h}]
Out[28]= {}
```

There are none. This verifies the *l* case of Theorem 8. *Mathematica* is, indeed, a powerful tool for reasoning about the probability calculus.[116]

[116]One can easily automate these techniques. In Fitelson (2001c), I present a *Mathematica* function **PrSAT** which, for any set *S* of inequalities, equations and inequations in (3-element) probability calculus, will determine whether *S* is satisfiable. If *S* is unsatisfiable, then **PrSAT** will say so, and if *S* is satisfiable, then **PrSAT** will output a probability model on which all of the statements in *S* are true.

Bibliography

Adams, E. (1998). *A Primer of Probability Logic*. Stanford, CA: CSLI Publications.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (Tsahkadsor, 1971), pp. 267–281. Budapest: Akadémiai Kiadó.

Carnap, R. (1945). On inductive logic. *Philosophy of Science 12*, 72–97.

Carnap, R. (1962). *Logical Foundations of Probability* (Second ed.). Chicago: University of Chicago Press.

Carnap, R. (1971). Inductive logic and rational decisions. In R. Jeffrey (Ed.), *Studies in Inductive Logic and Probability, Vol. I*, pp. 5–31. Berkeley: University of California Press.

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

Cox, R. T. (1961). *The Algebra of Probable Inference*. The Johns Hopkins Press, Baltimore, MD.

Csiszár, I. (1978). Information measures: a critical survey. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions and the Eighth European Meeting of Statisticians* (Tech. Univ. Prague, Prague, 1974)*, Vol. B*, pp. 73–86. Prague: Academia.

Dale, A. (1999). *A History of Inverse Probability: From Thomas Bayes to Karl Pearson* (Second ed.). New York: Springer-Verlag.

de Finetti, B. (1990). *Theory of Probability. Vols. 1 & 2*. Chichester: John Wiley & Sons Ltd. A critical introductory treatment, Translated from the Italian and with a preface by Antonio Machì and Adrian Smith, With a foreword by D. V. Lindley, Reprint of the 1975 translation.

Deák, I. (1990). *Random Number Generators and Simulation*. Budapest: Akadémiai Kiadó (Publishing House of the Hungarian Academy of Sciences). Translated and revised from the Hungarian by the author.

Earman, J. (1992). *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. Cambridge: MIT Press.

Eells, E. (1982). *Rational Decision and Causality*. Cambridge: Cambridge University Press.

Eells, E. (1983). Objective probability theory theory. *Synthese 57*(3), 387–442.

Eells, E. (1985). Problems of old evidence. *Pacific Philosophical Quarterly 66*, 283–302.

Eells, E. and B. Fitelson (2000). Measuring confirmation and evidence. *Journal of Philosophy XCVII*(12), 663–672.

Eells, E. and B. Fitelson (2001). Symmetries and asymmetries in evidential support. To appear in *Philosophical Studies*.

Efron, B. (1978). Controversies in the foundations of statistics. *American Mathematical Monthly 85*(4), 231–246.

Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician 40*(1), 1–11. With discussion and a reply by the author.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications. Vols. I & II* (Third ed.). New York: John Wiley & Sons.

Festa, R. (1999). Bayesian confirmation. In M. Galavotti and A. Pagnini (Eds.), *Experience, Reality, and Scientific Explanation*, pp. 55–87. Dordrecht: Kluwer Academic Publishers.

Fine, T. (1973). *Theories of Probability: An Examination of Foundations*. New York: Academic Press.

Fitelson, B. (1996). Wayne, Horwich, and evidential diversity. *Philosophy of Science 63*, 652–660.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science 66*, S362–S378.

Fitelson, B. (2001a). A Bayesian account of independent evidence with applications. Forthcoming in *Philosophy of Science*.

Fitelson, B. (2001b). Putting the irrelevance back into the problem of irrelevant conjunction. Unpublished manuscript.

Fitelson, B. (2001c). Using MATHEMATICA® to reason about the probability calculus. Unpublished manuscript.

Forster, M. (1994). Non-Bayesian foundations for statistical estimation, prediction, and the ravens example. *Erkenntnis 40*(3), 357–376.

Forster, M. (1995). Bayes and bust: Simplicity as a problem for a probabilist's approach to confirmation. *The British Journal for the Philosophy of Science 46*, 399–424.

Forster, M. (2000). Key concepts in model selection: Performance and generalizability. *Journal of Mathematical Psychology 44*, 205–231.

Forster, M. and I. Kieseppä (2001). The myth of reduction: Or why macro-probabilities average over counterfactual hidden variables. Forthcoming in *Philosophy of Science*.

Forster, M. and E. Sober (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science 45*(1), 1–35.

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

Gillies, D. (2000). *Philosophical Theories of Probability*. London: Routledge.

Glymour, C. (1980). *Theory and Evidence*. Cambridge, Mass: MIT Press.

Good, I. (1968). Corroboration, explanation, evolving probability, simplicity, and a sharpened razor. *The British Journal for the Philosophy of Science 19*, 123–143.

Good, I. (1983). *Good Thinking: The Foundations of Probability and its Applications*. Minneapolis: University of Minnesota Press.

Good, I. (1984). The best explicatum for weight of evidence. *Journal of Statistical Computation and Simulation 19*, 294–299.

Good, I. (1985). Weight of evidence: a brief survey. In *Bayesian Statistics, 2* (Valencia, 1983), pp. 249–269. Amsterdam: North-Holland. With discussion and a reply by the author.

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

Goosens, W. (1979). Alternative axiomatizations of elementary probability theory. *Notre Dame Journal of Formal Logic 20*(1), 227–239.

Guiaşu, S. (1977). *Information Theory with Applications.* New York: McGraw-Hill International Book Co.

Hacking, I. (1975). *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference.* London: Cambridge University Press.

Hailperin, T. (1996). *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications.* Lehigh University Press, Bethlehem, PA.

Hájek, A. (2001). What conditional probabilities could not be. Forthcoming in *Synthese.*

Halpern, J. (1999a). A counterexample to theorems of Cox and Fine. *Journal of Artificial Intelligence Research 10*, 67–85 (electronic).

Halpern, J. (1999b). Cox's theorem revisited. Technical addendum to: "A counterexample to theorems of Cox and Fine". *Journal of Artificial Intelligence Research 11*, 429–435 (electronic).

Heckerman, D. (1988). An axiomatic framework for belief updates. In L. Kanal and J. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 2*, pp. 11–22. New York: Elsevier Science Publishers.

Hellman, G. (1997). Bayes and beyond. *Philosophy of Science 64*(2), 191–221.

Hempel, C. (1945). Studies in the logic of confirmation. *Mind 54*, 1–26, 97–121.

Hempel, C. (1983). Kuhn and Salmon on rationality and theory choice. *The Journal of Philosophy 80*(10), 570–572.

Horvitz, E. and D. Heckerman (1986). The inconsistent use of certainty measures in artificial intelligence research. In L. Kanal and J. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 1*, pp. 137–151. New York: Elsevier Science Publishers.

Horwich, P. (1982). *Probability and Evidence*. Cambridge: Cambridge University Press.

Horwich, P. (1993). Wittgensteinian bayesianism. In *Midwest Studies in Philosophy Volume XVIII: Philosophy of Science*. University of Notre Dame Press.

Hosiasson-Lindenbaum, J. (1940). On confirmation. *Journal of Symbolic Logic 5*, 133–148.

Howson, C. and P. Urbach (1993). *Scientific Reasoning: The Bayesian Approach*. La Salle: Open Court.

Jeffrey, R. (1992). *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.

Jeffreys, H. (1998). *Theory of Probability*. New York: The Clarendon Press Oxford University Press. Reprint of the 1983 edition.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.

Kemeny, J. and P. Oppenheim (1952). Degrees of factual support. *Philosophy of Science 19*, 307–324.

Keynes, J. (1921). *A Treatise on Probability*. London: Macmillan.

Kolmogorov, A. (1956). *Foundations of Probability* (second english ed.). Providence, Rhode Island: AMS Chelsea Publishing.

Krantz, D., R. Luce, P. Suppes, and A. Tversky (1971). *Foundations of Measurement*, Volume 1. New York: Academic Press.

Kruse, M. (1997). Variation and the accuracy of predictions. *The British Journal for the Philosophy of Science 48*(2), 181–193.

Kruse, M. (1999). Beyond Bayesianism: comments on G. Hellman's "Bayes and beyond". *Philosophy of Science 66*(1), 165–174.

Kullback, S. (1997). *Information Theory and Statistics*. Mineola, NY: Dover Publications Inc. Reprint of the second (1968) edition.

Kullback, S. and R. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics 22*, 79–86.

Kyburg, H. (1983). Recent work in inductive logic. In T. Machan and K. Lucey (Eds.), *Recent Work in Philosophy*, pp. 87–150. Lanham: Rowman & Allanheld.

Levi, I. (1967). *Gambling with Truth*. Cambridge, Massachusetts: MIT Press.

Linhart, H. and W. Zucchini (1986). *Model Selection*. New York: John Wiley & Sons Inc.

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

Maher, P. (1993). *Betting on Theories*. Cambridge.

Maher, P. (1996). Subjective and objective confirmation. *Philosophy of Science 63*, 149–174.

Maher, P. (1997). Depragmatized Dutch book arguments. *Philosophy of Science 64*(2), 291–305.

Maher, P. (1999). Inductive logic and the ravens paradox. *Philosophy of Science 66*(1), 50–70.

Maher, P. (2000). Probabilities for two properties. *Erkenntnis 52*(1), 63–91.

Maher, P. (2001). Probabilities for multiple properties: The models of Hesse and Carnap and Kemeny. Forthcoming in *Erkenntnis*.

Maĭstrov, L. (1974). *Probability Theory: A Historical Sketch*. New York-London: Academic Press. Translated and edited by Samuel Kotz, *Probability and Mathematical Statistics, Vol. 23*.

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

Mortimer, H. (1988). *The Logic of Induction*. Paramus: Prentice Hall.

Myrvold, W. (1996). Bayesianism and diverse evidence: A reply to Andrew Wayne. *Philosophy of Science 63*, 661–665.

Niiniluoto, I. (1987). *Truthlikeness*. Dordrecht: Reidel.

Nozick, R. (1981). *Philosophical Explanations*. Cambridge: Harvard University Press.

Paris, J. (1994). *The Uncertain Reasoner's Companion: A Mathematical Perspective*. Cambridge: Cambridge University Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kauffman.

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

Pollard, S. (1999). Milne's measure of confirmation. *Analysis 59*, 335–337.

Popper, K. (1980). *The Logic of Scientific Discovery*. London: Routledge.

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

Ramsey, F. (1990). *Philosophical Papers*. Cambridge: Cambridge University Press. Edited and with a preface and an introduction by D. H. Mellor.

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

Reichenbach, H. (1956). *The Direction of Time*. University of California.

Reichenbach, H. (1971). *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability* (Second ed.). Berkeley: University of California Press. Translated from the German by Ernest H. Hutten and Maria Reichenbach.

Robert, C. (1994). *The Bayesian Choice: A Decision-Theoretic Motivation*. New York: Springer-Verlag. Translated and revised from the French original by the author.

Roeper, P. and H. Leblanc (1999). *Probability Theory and Probability Logic*. Toronto: University of Toronto Press.

Rosenkrantz, R. (1977). *Inference, Method and Decision*. Dordrecht: D. Reidel.

Rosenkrantz, R. (1981). *Foundations and Applications of Inductive Probability*. Atascadero, Calif.: Ridgeview.

Rosenkrantz, R. (1994). Bayesian confirmation: Paradise regained. *The British Journal for the Philosophy of Science 45*, 467–476.

Ross, S. (1994). *A First Course in Probability* (fourth ed.). Englewood Cliffs, NJ: Prentice Hall.

Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.

Savage, L. (1972). *The Foundations of Statistics* (second revised ed.). New York: Dover Publications.

Schervish, M. (1995). *Theory of Statistics*. New York: Springer-Verlag.

Schlesinger, G. (1995). Measuring degrees of confirmation. *Analysis 55*, 208–212.

Schum, D. (1994). *The Evidential Foundations of Probabilistic Reasoning*. New York: John Wiley & Sons.

Seidenfeld, T. (1979). Why I am not an objective Bayesian; some reflections prompted by Rosenkrantz. *Theory and Decision 11*(4), 413–440.

Skyrms, B. (1984). *Pragmatics and Empiricism*. New Haven: Yale University Press.

Skyrms, B. (1992). Review of papers by Glymour, Grimes, and Waters on confirmation theory. *Journal of Symbolic Logic 57*(2), 756–758.

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

Sober, E. (1994a). Contrastive empiricism. In *From A Biological Point of View: Essays in Evolutionary Philosophy*, pp. 114–135. Cambridge: Cambridge University Press.

Sober, E. (1994b). No model, no inference: A bayesian primer on the grue problem. In D. Stalker (Ed.), *Grue! The New Riddle of Induction*. Chicago: Open Court.

Sober, E. (1994c). The principle of the common cause. In *From A Biological Point of View: Essays in Evolutionary Philosophy*, pp. 158–174. Cambridge: Cambridge University Press.

Stigler, S. (1990). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge, MA: The Belknap Press of Harvard University Press. Reprint of the 1986 original.

Swinburne, R. (1973). *An Introduction to Confirmation Theory*. London: Methuen.

Székely, G. (1986). *Paradoxes in Probability Theory and Mathematical Statistics*. Dordrecht: D. Reidel. Translated from the Hungarian by Márta Alpár and Éva Unger.

van Fraassen, B. (1982). *The Scientific Image*. New York: Oxford University Press.

van Fraassen, B. (1989). *Laws and Symmetry*. New York: Oxford University Press.

Wayne, A. (1995). Bayesianism and diverse evidence. *Philosophy of Science 62*, 111–121.

Wittgenstein, L. (1953). *Philosophical Investigations*. New York: The Macmillan Co. Translated by G. E. M. Anscombe.

Wolfram, S. (1999). *The MATHEMATICA® Book* (Fourth ed.). Wolfram Media, Inc., Champaign, IL.