4.3 Evidential Relevance

Arthur Merin

Summary

Signed evidential relevance among truth-valuable propositions has qualitative, comparative, and quantitative aspects. We explore the abstract theory of qualitative relevance quite fully, notably so the role of conditional irrelevance. Regarding ordinal and quantitative aspects of relevance, we offer a notational proposal and two extensions of the state of the art. First, we demonstrate task-related conditions for ordinal equivalence of all relevance measures satisfying a highly intuitive desideratum shared by the major relevance measures. This affords a useful degree of measure independence for many comparative relevance judgments. Second, we elicit relevance properties of disjunctive evidence to complement familiar facts about conjunctive evidence. The central role of relevance for reasoning is emphasized throughout.

1. Intuitions of Evidence in Argumentation

Will some proposition or statement, A, speak for the truth of another proposition, B, or against it? Questions of this kind are ubiquitous, and people answer them within the context of their general beliefs about the world. Example: Let A = "There are dark clouds in the sky today," B_1 = "It will rain today," B_2 = "It will be dry today," B_3 = "The Dow Jones will rise today." In typical belief contexts, A will speak for B_1, against B_2, and neither for nor against B_3. Similar relations obtain between statements about past events, as well as between possible experimental results and timeless scientific hypotheses.

Here is a way to reconstruct this kind of reasoning. We are uncertain about the truth values, True or False, of B_1, B_2, and B_3. We may be uncertain about the truth value of A. And we can distinguish, in principle, at least a handful of degrees of credence between certainty of truth and certainty of falsity. Now comes the principal ingredient of the reconstruction. If, being uncertain about A ("dark clouds"), we learnt for certain that A is true—just this much!—then our degrees of credence should move upward for B_1 ("rain"), go downward for B_2 ("dry"), and remain unchanged for B_3 ("Dow up"). We need not actually become certain that B_1 is true. Neither need we become certain that B_2 is false. These predicaments are but special cases of "speaking for" and "speaking against."

Next suppose that our beliefs are as before, but that we learn that A ("dark clouds") is false; equivalently, $\neg A$ ("no dark clouds") is true. The suggestion here is as follows. Our credence in B_3 ("Dow up") would again remain unaffected. Our degree of credence for B_2 ("no rain") would go up. And our credence in B_1 ("rain") would go down. If so, learning that A is false must afford a basis for determining what A speaks for, against, and neither way. A will speak for B if $\neg A$ speaks against B.

In the diction of jurisprudence and of science, "speaking for" B means being *evidence for B*. "Speaking against" B means being *evidence against B*. In this evidential sense, we say that A is *positively relevant* ("*positive*") to B in a credal context or state *Cred* if and only if (iff) it speaks for B in *Cred*, *negatively relevant* ("*negative*") to B iff it speaks against B in *Cred*, evidentially *irrelevant* to B if it speaks neither for nor against B in *Cred*, and *relevant* iff not irrelevant.

Plausible as the idea may appear, it still remains vague and prone to apparent exceptions. Some framework for specifying it is needed to give it substance.

2. Evidential Relevance Explicated in Probability Theory

To turn intuitions about relevance into a theory, one interprets degrees of credence in A, B, and so on in a credal state, as credal probability values $P(A)$, $P(B)$, and so forth.[1]

Changes of credal state, from *Cred* to *Cred′*, are then represented by changes of probability measures, from P to P'. Any such change will yield credal reassessments $P'(A)$, $P'(B)$, and so on. However, it is one particular belief change regime that yields a workable explication of the intuitive notion of relevance. This regime is intimately associated with the concept of conditional probability, in terms of which relevance is standardly definable.

Definition 1. The *conditional probability* of A given B, notated $P(B|A)$, is defined by $P(B|A) =_{df} P(AB)/P(A)$ whenever $P(A) \neq 0$, else undefined.[2]

We gloss it, roughly for now, as the probability that B would have if A were assumed. Recall now section 1. Suppose we learn A, no more, no less, and without any doubts or reservations. Then we should be in a new, "posterior" credal state, P', validating $P'(A) = 1$; $P'(\neg A) = 0$. Let our update be by *conditioning P on A*. Then the posterior state P' may be written P_A, and for all B, the posterior probability of B equals its conditional probability: $P_A(B) = P(B|A)$.

The probability-change explication of speaking "for" or "against" makes A positive to B in P iff $P_A(B) > P(B)$ and negative to B iff $P_A(B) < P(B)$. This yields

Definition 2. A is $\left\{\begin{matrix} \textit{positive} \\ \textit{irrelevant} \\ \textit{negative} \end{matrix}\right\}$ to B (in P) iff

$$\left\{\begin{matrix} P(B|A) > P(B) \\ P(B|A) = P(B) \text{ or } P(A) = 0 \\ P(B|A) < P(B) \end{matrix}\right\}.$$

Multiplying the (in)equalities of definition 2 with $P(A)$ yields the equivalent definition 3, which we state with obvious abbreviations and using the standard symbol $(A \perp B)_P$ for "A is irrelevant to B in P":

$$\textit{Definition 3.}\ A \left\{\begin{matrix} pos \\ \perp \\ neg \end{matrix}\right\}_P B \text{ iff } P(AB) \gtreqless P(A) \cdot P(B).$$

Definition 3 depends on definition 2 for the underlying intuition. In return, it needs no completion clause and reveals the symmetry of relevance relations.

3. Properties of Qualitative Relevance and Extreme Positivity

Qualitative relevance of A to B, and of B to A, is specified only in terms of polarity. I present six of its properties in order of intuitive simplicity or importance.

(1) *Symmetry of qualitative relevance*: Explicated in probability theory, positive relevance, negative relevance, and irrelevance are *symmetric relations*:

$$\textit{Theorem 1.}\ A \left\{\begin{matrix} pos \\ \perp \\ neg \end{matrix}\right\}_P B \text{ iff } B \left\{\begin{matrix} pos \\ \perp \\ neg \end{matrix}\right\}_P A.$$

The proof is obvious from definition 3. For instance, if A speaks for B, then B speaks for A, however weakly (intuitively).

(2) *Polarity reversal by negation*: Negating just one of A and B reverses relevance polarity and preserves irrelevance. Negating both of A and B leaves relevance polarity unchanged:

$$\textit{Theorem 2.}\ A \left\{\begin{matrix} pos \\ \perp \\ neg \end{matrix}\right\}_P B \text{ iff } \neg A \left\{\begin{matrix} neg \\ \perp \\ pos \end{matrix}\right\}_P B \text{ iff } A \left\{\begin{matrix} neg \\ \perp \\ pos \end{matrix}\right\}_P \neg B \text{ iff } \neg A \left\{\begin{matrix} pos \\ \perp \\ neg \end{matrix}\right\}_P \neg B.$$

(3) *Certainty implies irrelevance*: Uncertainty is a necessary condition for relevance:

Theorem 3. A proposition A that has extreme probability, $P(A) = 0$ or $= 1$, is irrelevant, under P, to any other proposition and to itself.

For proof instantiate definition 3. To be relevant (or for anything to be relevant to it) in P, a proposition A must therefore be P-contingent, $0 < P(A) < 1$.

(4) *Extreme positive and negative relevance*: Classical logical entailment, $A \vdash B$, is transitive: $A \vdash B$ and $B \vdash C$ together imply $A \vdash C$. Positive relevance is *not* in general transitive. That is, $A\ pos_P\ B$ and $B\ pos_P\ C$ need not imply $A\ pos_P\ C$. An example will be given shortly. However, there are conditions under which positive relevance *is* transitive. One special case thereof engages "conclusive-evidence-for" as explicated in definition 4a:

Definition 4 (Carnap 1950, modified). (a) A is *extremely positive* to B iff $P(B) < P(B|A) = 1$. (b) A is *extremely negative* to B iff extremely positive to $\neg B$. We abbreviate (a) as $A\ expos_P\ B$ and (b) as $A\ exneg_P\ B$.

Clause (a) succinctly describes in a credal probability theory what is conveyed when someone asserts, without reservations, an indicative conditional "If A, then B." Of course, such an assertion can also be made with less certainty. Extreme positivity explicates what colloquial language mostly means by "entailment." The idea is that A is a compelling and, in this sense, *sufficient reason* for B. It is *sufficient* by $P(B|A) = 1$, which makes B a conditional certainty. It is a *reason* in virtue of $P(B) < P(B|A)$. Reasonhood has a reflex in our sense of evidential aesthetics: "If the Dow rises (=A), (then) 7 is a prime number (=B)" sounds nonsensical even when $P(A) > 0$ is uncontroversial, as normally it is, and $P(B|A) = 1$ must therefore hold because $P(B) = 1$ holds. Classical entailment, $A \vdash B$, is related to extreme relevance by

Lemma 1. If $0 < P(A), P(B) < 1$, and A entails B ($A \vdash B$), then $A\ expos_P\ B$.

Entailment, $\vdash$, is a transitive relation. So is its probabilistic counterpart, *P-entailment*, defined by $A \vdash_P B =_{df} P(A \rightarrow B) = 1$ (where $\rightarrow$ is material implication). Extreme positivity inherits the transitivity of $\vdash$ and $\vdash_P$:

Theorem 4. Extreme positive relevance is transitive: if A *expos*$_P$ B and B *expos*$_P$ C, then A *expos*$_P$ C.

(5) *Conditional irrelevance and conditional transitivity of positive relevance*: When positive relevance of A to B and of B to C is short of extreme, transitivity may fail. Here is an example: Assume A *expos*$_P$ B, C *expos*$_P$ B, and $P(AC) = 0$. These assumptions are easily satisfiable. For example, A = "Trump has won," B = "The U.S. has a new president," and C = "Clinton has won." Then A *pos*$_P$ B, B *pos*$_P$ C (because of symmetry), yet A *exneg*$_P$ C, hence A *neg*$_P$ C.

Transitivity is, however, ensured by a supremely useful property. Relevance relations can be *conditional* on a proposition. On noting that $P(A) = P(A \mid \top)$ is a probability theorem (where $\top$ is any tautology), definition 3 becomes a special case of

Definition 5. A is *irrelevant* to (*independent* of) C *conditionally* on B under P, $(A \perp C \mid B)_P$, iff $P(AC \mid B) = P(A \mid B) \cdot P(C \mid B)$. When each of $(A \perp C \mid B)_P$ and $(A \perp C \mid \neg B)_P$ holds, we write $(A \perp C \mid \pm B)_P$.

We now have a sufficient condition for transitivity of positive relevance that need not be extreme:

Theorem 5. If A and B are irrelevant to one another conditionally on B and on $\neg B$, that is, if $(A \perp C \mid \pm B)_P$, then mutual positivity of A and B and mutual positivity of B and C imply mutual positivity of A and C.[3]

Condition $(A \perp C \mid \pm B)_P$ becomes more intuitable on realizing that the extreme-relevance configuration of A *expos*$_P$ B and B *expos*$_P$ C entails the antecedent of the following lemma and thus $(A \perp C \mid \pm B)_P$:

Lemma 2. If $P(B \mid A) = 1$, $P(C \mid B) = 1$, and $P(B) < 1$, then $(A \perp C \mid \pm B)_P$.

Hence, theorem 5 properly generalizes intuition-friendly theorem 4.

(6) *Conditional irrelevance and paradoxa of relevance*: Theorem 5 shows that $(A \perp C \mid \pm B)_P$ does not imply $(A \perp C)_P$. This counterintuitive fact constitutes an instance of a well-known statistics paradox. Suppose **B** and **¬B** specify complementary properties of members of a select population, say, "British" and "non-British." Let **A** stand for having been administered pharmaceutical **Φ**, and **C** for being cured of disease Δ. These properties and their complexes can be turned into propositions ($A, B, \neg B, C$, etc.), whose credal probabilities reflect statistical frequencies. (Think of proposition D as "ι has property **D**," where ι is a randomly sampled individual.) Now suppose we investigate the relevance of A to C. Then theorem 5 implies that A may be positively relevant to C in the population as a whole, but irrelevant to C in each of subpopulations **B** and **¬B** (because, e.g., the drug **Φ** is actually inefficacious concerning Δ, but both the drug administration and the spontaneous remission rate are higher among British than among non-British citizens). This is an instance of "Simpson's Paradox" (see chapter 7.1 by Pearl, this handbook).

In return, issue-conditional independence $(A \perp C \mid \pm B)_P$ (referring to the issue whether B or $\neg B$) precludes otherwise possible, intuitively paradoxical relevance configurations such as the one where A and C are positive to B, but AC is irrelevant or negative to B (cf. Carnap, 1950, chapter 6). For Reichenbach (1956) shows, as a special case of a result for various combinations of relevance polarities:

Theorem 6. Suppose that $(A \perp C \mid \pm B)_P$ holds, A and B are mutually positive, and so are B and C. Then B and the conjunction AC are mutually positive.

Reichenbach's intended interpretations for theorems 5 and 6 had B as a *common cause* of A and C. Pearl (1988; see chapter 7.1 by Pearl, this handbook) surmised that medical diagnostic categories B (diseases, syndromes) are formed to make symptoms A and C conditionally independent given $\pm B$, by identifying B as their common cause.[4] A test of Pearl's thesis would thus be to search for diseases that exhibit Carnap-paradoxical symptomatologies. Anyone's credence in the thesis should increase if none were found. But why? Simply speaking, because finding a disease with a paradoxical symptomatology would be extremely negative to the thesis, which implies that finding none is positive to it. We now address this idea more generally.

4. An Application of Qualitative Relevance Reasoning

The most familiar applications of theorem 1 (symmetry) and theorem 2 (polarity reversal under negation) presuppose a sorting of propositions into two types: (1) "hypothesis" statements or "theories," labeled "H," and (2) observational "data reports" or "evidence statements," labeled "D" or "E." Suppose a hypothesis H having nonzero probability entails some data report D, which might be one possible outcome of an intended, future investigation. In symbols: $H \vdash D$. Then $P(D \mid H) = 1$ holds. Hence, provided $0 < P(D)$, $P(H) < 1$ holds, H is extremely positive, and thus positive, to D. By symmetry, $P(H \mid D) > P(H)$ must also hold. Suppose outcome D is then observed and P conditioned on it. Then $P_D(H) > P(H)$ must hold. H has become *more credible*. Alternatively, suppose outcome $\neg D$ is observed. By theorem 2, $P_{\neg D}(H) < P(H)$ holds, so H becomes *less credible*. This is what the relevance laws developed so far entail.

Complications arise when (a) D is already-known, "old" evidence to be explained, rather than predicted, and (b) P is required to reflect this knowledge. By theorem 3, such D must be P-irrelevant: $P_D(H) = P(H)$ (Glymour, 1980). A standard example has H instantiated by Einstein's theory of general relativity, which entailed D, a long-known and seemingly anomalous fact about the orbit of planet Mercury. This relationship was widely judged to increase confidence in H. Reconstructions of this judgment appeal to hypothetical, partly nonconditioning update histories (Jeffrey, 2004), introduce probabilities over primitive propositions about entailment or support (cf. Earman, 1992; Fitelson & Hartmann, 2015), or leave D slightly in doubt and choose a relevance measure (see below) that gives near-certain D an appreciable amount of relevance (cf. Hájek & Joyce, 2008). Working statisticians have, in this respect, an easier life. Faced with known data, they usually draw on ready-made probability models for conditional statistical frequencies of occurrence. This replaces commitment (b) by appropriate probability distributions for hypothetical predictions under repeatable conditions.[5]

5. Comparative Strength and Amount of Relevance

Consider propositions A_1 = "There are dark clouds," A_2 = "The weather forecast predicts rain," A_3 = "The weather forecast predicts dry conditions," and B = "It will rain." Imagine, as seems natural, that each of A_1 and A_2 would speak for B, while A_3 speaks against B. Now suppose that you wonder which of A_1 and A_2, if any, would speak more strongly for B than the other does. Or imagine that you learn A_1 and A_3 and wonder if A_1A_3 speaks for B, or against B, or does neither. Or imagine asking which of B ("It will rain") and B' ("It will be snowing") is more strongly spoken for by A_1 ("dark clouds") in a given credal state. Neither of these questions can usually be resolved by purely qualitative considerations.

Suppose, then, that A_1 and A_2 are each positive to B in credal state P. If we compare A_1 and A_2 by strength of relevance to B, the resultant ordering within the probabilistic explication will, in general, depend on measurements of strength. This fact already suffices to make measures of relevance a pertinent topic for anyone interested in probabilistic relevance. Rankings may indeed depend on our choice of measuring function. This is much as could happen in ranking companies by asset value: rankings may differ according to the accounting convention used.

A further problem arises. If A and B are mutually relevant, A may be more strongly relevant to B than B is to A under one measure, less relevant under another, and equally relevant under a third. A helpful measurement notation should thus make the intended direction immediately clear, without tacitly relying on the classification of content as, say, evidence or hypothesis, which is often application dependent (see below). What motivates conventional assignments of relevance direction may then be freely investigated.

5.1 Notation and Terminology

In recent literature, degree of relevance is widely glossed as "*amount of* '*(incremental) confirmation*' of '*hypothesis*' H provided by '*evidence*' E under *probability measure* P and *measure of type* C." Notation is often "$C_P(H, E)$," sometimes "$C_P(E, H)$." The $C_P(\cdot, \cdot)$ format affords no natural mnemonic of direction. True, its gloss will suggest a direction: *from* "evidence" *to* "hypothesis." Yet often enough we reason from hypotheses to possible data or find no warrant for the evidence/hypothesis type distinction. Thus, reliance on E and H abets premature identification of *structural argument position* with *content type*. Concomitantly, labels such as "confirmation," "corroboration," and "support" need a prefix "incremental" to avoid being misunderstood as designating conditional probability, $P(H | E)$. They also sound awkward with the prefix "negative." "Relevance" (Keynes, 1921; Carnap, 1950; Jeffrey, 2004) has no such problems.

All this speaks for a content-neutral approach to symbolization, at least alongside its common philosophy-of-science instantiation. I propose an iconic, signpost-like separator symbol for proposition arguments that indicates conventional measurement direction as the wording does in "relevance of A to B."

Definition 6 (Notation, Terminology). Let $R_{\mathsf{T}}(A \succ B)_P$ designate—subject to conditions specified in definition 8—the *numerical amount of relevance* of proposition A to proposition B as assessed in terms of a relevance measure of type **T** under a credal probability measure P. Refer to the proposition arguments A and B preceding and following the signpost separator $\succ$ as being respectively in *source position* and in *target position*.

Source $\succ$ target direction affects numerical measurement values, at some point or other, for all relevance measures that do *not* have the property introduced in

Definition 7. A relevance measure $R_T(A \succ B)_P$ is *amount-symmetric* iff $R_T(B \succ A)_P = R_T(A \succ B)_P$ for all A, B, P.

Amount-symmetric R_T are, in this unambiguous sense, *nondirectional*.

5.2 Two Definitional Conditions on All Relevance Measures

Many distinct measures of relevance have been discussed in the literature (see, e.g., Eells & Fitelson, 2002; Crupi & Tentori, 2016). All measures validate

Definition 8. A function $R_T(A \succ B)_P$ mapping pairs $\langle P, \langle A, B\rangle\rangle$ consisting of a probability measure P and a pair $\langle A, B\rangle$ of propositions in the domain of P to numeric values is a *probabilistic relevance measure* iff it satisfies

(a) *Transparency*: $R_T(A \succ B)_P$ must be a smooth (i.e., differentiable) function of the probability distribution P over the atoms AB, $A \wedge \neg B$, $\neg A \wedge B$, $\neg A \wedge \neg B$ of the propositional algebra generated by A and B, and its value range must be a subinterval, proper or improper, of the extended real numbers $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, \infty\}$.[6]

(b) *Quality Representation*: $R_T(A \succ B)_P$ must represent qualitative relevance polarity: its value range must contain an element $\iota := \iota_T$ such that

$$R_T(A \succ B)_P \gtreqless \iota_T \text{ iff } A \left\{\begin{matrix} pos \\ \perp \\ neg \end{matrix}\right\}_P B.$$

Measures in actual use have for neutral value either $\iota = 0$, the "neutral" identity element for addition, or $\iota = 1$, the multiplicative identity. Logarithmic transformations map multiplication to addition and thus multiplicative to additive identities.

In what follows, we mostly adopt the common format $R(E \succ H)$, which identifies the "source" proposition with "evidence" E and the "target" proposition with a "hypothesis" H. This makes for easy cross-reference to the confirmation literature, just as A and B do to the content-neutral mathematical literature.

5.3 The Big Three Measures of First Resort and Their Ordinal Equivalents

The "Big Three"—**D**, **Q**, and **L**—are the oldest and most commonly used relevance measures. Their formal definitions are given below. (**D**, **Q**, and **L** stand for "Difference," "Quotient," and "Likelihood.") Also given are the most useful and most common transforms for **Q** and **L**, namely, **QL** and **LL**.

$$\textbf{D:}\quad R_D(E \succ H)_P =_{df} \begin{cases} P(H|E) - P(H), & \text{if } P(E) \neq 0, \\ 0, & \text{else.} \end{cases}$$

$$\textbf{Q:}\quad R_Q(E \succ H)_P =_{df} \begin{cases} \dfrac{P(H|E)}{P(H)} = \dfrac{P(EH)}{P(E)P(H)} = \dfrac{P(E|H)}{P(E)}, & \text{if } P(E), P(H) \neq 0, \\ 1, & \text{else.} \end{cases}$$

$$\textbf{L:}\quad R_L(E \succ H)_P =_{df} \begin{cases} \dfrac{P(E|H)}{P(E|\neg H)}, & \text{if } P(H), P(\neg H), P(E|\neg H) \neq 0, \\ \infty, & \text{if } P(H), P(\neg H) \neq 0 \text{ and } P(E|\neg H) = 0 \neq P(E|H), \\ 1, & \text{else.} \end{cases}$$

$$\textbf{QL:}\quad R_{QL}(E \succ H)_P =_{df} \log R_Q(E \succ H)_P = \log P(EH) - \log P(E) - \log P(H).$$

$$\textbf{LL:}\quad R_{LL}(E \succ H)_P =_{df} \begin{cases} \log R_L(E \succ H)_P = \log P(E|H) - \log P(E|\neg H), & \text{if } P(H), P(\neg H), P(E|H), P(E|\neg H) \neq 0, \\ \infty, & \text{if } P(E|\neg H) = 0 \neq P(E|H), \\ -\infty, & \text{if } P(E|H) = 0 \neq P(E|\neg H), \\ 0, & \text{else.} \end{cases}$$

Let me introduce two further measures to be mentioned below: **KO** and **Z**. Formal definitions for these are given at the bottom of this page.

Even more measures are discussed in the literature.

The "difference measure" R_{D} (Hugh MacColl's δ of 1897 or earlier; cf. Keynes, 1921, pp. 154–155) has irrelevance value $\iota = 0$ and range $(-1, 1)$. The "coefficient of influence" (W. E. Johnson, 1921 or earlier, published in Keynes, 1921) or "relevance quotient" **Q**[8] has $\iota = 1$ and range $[0, \infty)$. Its variant expression $\dfrac{P(EH)}{P(E)P(H)}$ reveals its amount symmetry: $R_{\mathsf{Q}}(E \succ H)_P = R_{\mathsf{Q}}(H \succ E)_P$. The "likelihood ratio" **L** has $\iota = 1$ and range $[0, \infty]$. Modulo logarithms, **L** was commended by C. S. Peirce, A. M. Turing, and I. J. Good. It has an information-theoretic claim to primacy, being both (i) sufficient and (ii) necessary to capture all the information in the data E that bears on the hypothesis bipartition $\{H, \neg H\}$. It conveys all of it (i) and any other sufficient statistic factors through it (ii) (cf. Savage, 1954, pp. 128–140).

The logarithmic **QL**-transform (Good, 1950) has $\iota = 0$ and range $[-\infty, \infty)$. The "log-likelihood ratio" **LL** has $\iota = 0$ and range $[-\infty, \infty]$ and is widely used in statistics. The "sigmoid" transform **KO** (Kemeny & Oppenheim, 1952) of **LL** has $\iota = 0$ and squashes the range to $[-1, 1]$. Finally, for the measure **Z** (Crupi & Tentori, 2013) the irrelevance value ι is 0 and the range is $[-1, 1]$. The various transforms' existence suffices to motivate a highly general concept.

Definition 9. Measure types **T** and **T**′ are *ordinally equivalent* ($\mathbf{T} \sim_{\uparrow} \mathbf{T}'$) iff, for all possible P and propositions A, B, C, D in the domain of P, $R_{\mathsf{T}}(A \succ B)_P \gtreqless R_{\mathsf{T}}(C \succ D)_P$ iff $R_{\mathsf{T}'}(A \succ B)_P \gtreqless R_{\mathsf{T}'}(C \succ D)_P$. Equivalently, $\mathbf{T} \sim_{\uparrow} \mathbf{T}'$ iff there exists some strictly monotonically increasing (isotonic) function f for which $\mathbf{T}' = f(\mathbf{T})$.[9]

Distinct, but ordinally equivalent, measures may nonetheless differ in convenience of range and, more important for cognitive-engineering applications, in various algebraic properties.

The Big Three have appealing interpretations as additive or multiplicative operators. They map the *prior* credal value of a "target" proposition H to its value in the *posterior* credal state obtained by conditioning the prior P on the "source" proposition E. For **D** and **Q**, the credal values are probabilities. For **L** they are odds, $O(H)$ and $O(H \mid E)$, from which probabilities may be recomputed.[10]

$$P(H \mid E) = P(H) + R_{\mathsf{D}}(E \succ H)_P, \tag{1}$$
$$P(H \mid E) = P(H) \cdot R_{\mathsf{Q}}(E \succ H)_P, \tag{2}$$
$$\log P(H \mid E) = \log P(H) + R_{\mathsf{QL}}(E \succ H)_P, \tag{2L}$$
$$O(H \mid E) = O(H) \cdot R_{\mathsf{L}}(E \succ H)_P, \tag{3}$$
$$\log O(H \mid E) = \log O(H) + R_{\mathsf{LL}}(E \succ H)_P. \tag{3L}$$

Observe that equations (2) and (3) express, respectively,

Bayes' Theorem: $P(H|E) = \dfrac{P(E|H)P(H)}{P(E)}$,

Bayes' Theorem in odds form:

$$\frac{P(H|E)}{P(\neg H|E)} = \frac{P(H)}{P(\neg H)} \cdot \frac{P(E|H)}{P(E|\neg H)}.$$

5.4 Axiomatic Characterization

A most illuminating characterization of the Big Three—although only up to ordinal equivalence—is given by Crupi, Chater, and Tentori (2013) and Crupi and Tentori (2016). Modulo labeling, they consider four axioms for measure types **T**:

Posterior Probability Decisiveness (POP-D): For all H, E_1, E_2, P: $R_{\mathsf{T}}(E_1 \succ H)_P \gtreqless R_{\mathsf{T}}(E_2 \succ H)_P$ iff $P(H|E_1) \gtreqless P(H|E_2)$.

Likelihood Decisiveness (LIK-D): For all E, H_1, H_2, P: $R_{\mathsf{T}}(E \succ H_1)_P \gtreqless R_{\mathsf{T}}(E \succ H_2)_P$ iff $P(E|H_1) \gtreqless P(E|H_2)$.

Modularity for Conditionally Independent Evidence (MCIE): For all H, E_1, E_2, P: $R_{\mathsf{T}}(E_2 \succ H)_{P_{E_1}} = R_{\mathsf{T}}(E_2 \succ H)_P$ whenever $(E_1 \perp E_2 \mid \pm H)_P$, where, as before, $P_E =_{\mathrm{df}} P(\cdot|E)$, that is, P conditioned on E.

KO: $R_{\mathsf{KO}}(E \succ H)_P$

$$=_{\mathrm{df}} \begin{cases} \dfrac{P(E \mid H) - P(E \mid \neg H)}{P(E \mid H) + P(E \mid \neg H)} = \dfrac{P(EH) - P(E)P(H)}{P(H)P(E\neg H) + P(\neg H)P(EH)}, & \text{if } P(E), P(\neg E), P(H), P(\neg H) \neq 0, \\ 0, & \text{else.}^{7} \end{cases}$$

Z:
$$R_{\mathsf{Z}}(E \succ H)_P =_{\mathrm{df}} \begin{cases} \dfrac{P(H|E) - P(H)}{1 - P(H)}, & \text{if } P(E) \neq 0, P(H) \neq 1, \text{and } P(H|E) > P(H), \\ \dfrac{P(H|E) - P(H)}{P(H)}, & \text{if } P(E), P(H) \neq 0 \text{ and } P(H|E) \leq P(H), \\ 0, & \text{else.} \end{cases}$$

Modularity for Disjoint Hypothesis Disjuncts (MDHD): For all E, H_1, H_2, P: if $P(H_1H_2) = 0$, then $R_\mathsf{T}(E \succ H_1 \vee H_2)_P \gtreqless R_\mathsf{T}(E \succ H_1)_P$ iff $P(H_2 \mid E) \gtreqless P(H_2)$.[11]

Then they derive a representation theorem:

Theorem 7 (Crupi et al., 2013; Crupi & Tentori, 2016). Let R_T be a real-valued function satisfying Transparency (definition 8a) with the possible exception of its smoothness requirement. Then R_T is identical to, or is an isotonic transformation of,

(i) R_D iff it validates POP-D and MDHD,
(ii) R_Q iff it validates POP-D and LIK-D,
(iii) R_L iff it validates POP-D and MCIE.

The challenging "if"-parts of theorem 7 yield ordinal equivalence classes $[\mathbf{T}] =_{df} \{\mathbf{T}' \mid \mathbf{T}' \sim_\uparrow \mathbf{T}\}$ of measures **T**.[12] To determine specific measures **D**, **Q**, **L**, **QL**, **LL**, **KO**, and **Z**, further conditions are needed, among them "smoothness."

We now discuss the four axioms. Note first and importantly that POP-D and MCIE legislate for measurements comparing relevance of two "source/evidence" propositions E_1, E_2 to a given "target/hypothesis" H, while LIK-D and MDHD apply to comparisons of hypotheses H_1, H_2 for given E. Thus, POP-D and LIK-D do not compete and are indeed each validated by **Q**.

POP-D seems mandatory to most commentators. Its denial feels counterintuitive, and POP-D has also been argued to be necessary for enumerative induction based on the proportion of favorable and unfavorable outcomes of a series of structurally identical experiments (Steel, 2003). Only measures validating POP-D should thus be attractive as measures of first resort.[13]

LIK-D is often dubbed the "Law of Likelihood."[14] LIK-D feels less compelling than POP-D. However, intuitions on single axioms have limited normative value. Users who like POP-D and find direction-assignments unwarranted will adopt LIK-D, because POP-D and amount symmetry entail LIK-D. Users who value POP-D and prefer nonsymmetry at least for extreme relevance must allow violations of LIK-D.

MCIE, found in Good (1960) and Heckerman (1988), has its much desired special case in the respective conditional multiplicativity of **L** and additivity of **LL**. For, we have

Lemma 3: If $(E_1 \perp E_2 \mid \pm H)$ then $R_\mathsf{L}(E_1E_2 \succ H)_P = R_\mathsf{L}(E_1 \succ H)_P \cdot R_\mathsf{L}(E_2 \succ H)_P$ and thus $R_\mathsf{LL}(E_1E_2 \succ H)_P = R_\mathsf{L}(E_1 \succ H)_P + R_\mathsf{LL}(E_2 \succ H)_P$.

These practical benefits of MCIE are unavailable for other isotonic transforms of **L** (e.g., **KO**).

MDHD looks sensible but faces an embarrassment (cf. Crupi & Tentori, 2016). Suppose E is extremely negative to H. Now weaken H to $H' = H \vee G$, by a disjunct G that is P-irrelevant to E and P-disjoint from H, that is, $P(GH) = 0$. Then $R_\mathsf{D}(E \succ H)_P = R_\mathsf{D}(E \succ H')_P$, even though E speaks conclusively against H but only inconclusively against H'. Since the **L**- and **Q**-measures do not face this problem, the property illustrates ordinal inequivalence of **D** to each of them.

5.5 Measure Dependence and Taskwise Independence

To what extent are purely ordinal assessments of comparative relevance measure dependent? The Big Three, **D**, **Q**, **L**, are known to be pairwise ordinally inequivalent. To localize the inequivalences, I offer the following results.

Theorem 8. For the task of ranking multiple "target" propositions H_j ($j = 1, \ldots, m \geq 2$) by strength $R(E \succ H_j)_P$ of relevance of a given "source" proposition E to each of them, measures **D**, **Q**, and **L** are pairwise ordinally inequivalent.

Theorem 9. For the task of ranking multiple "source" propositions E_i ($i = 1, \ldots, n \geq 2$) by strength $R(E_i \succ H)_P$ of relevance to a given "target" proposition H, all relevance measures validating POP-D (e.g., **D**, **Q**, **L**, **QL**, **LL**, **Z**) are pairwise ordinally equivalent.

Theorem 9 holds because POP-D prescribes an ordinal assessment schedule for any fixed target H and variable pair of source propositions $E_i, E_{i'}$. Indeed, theorem 9 and the nondirectionality of amount-symmetric **Q** (section 5.3) entail

Lemma 4. If POP-D is a soundness requirement on measures $R_\mathsf{T}(E \succ H)_P$, then sound ordinal ranking of "source" propositions E_i ($i = 1, \ldots, n \geq 2$) by strength of relevance to a given "target" proposition H does not depend on the direction of measurement.

This lends credibility to the probabilistic explication of many judgments of comparative relevance that might otherwise seem critically measure dependent. Applications include solutions to Hempel's "Raven Paradox" (Hosiasson-Lindenbaum, 1940; more accessibly in Earman, 1992; Hájek & Joyce, 2008; Crupi, 2015). In abstract outline, this is the following predicament:

(1) H ("All ravens are black") is supported by the conjunctive proposition $A = A_1A_2$ (where A_1 = "a is a raven" and A_2 = "a is black"). Here "a" stands for a randomly encountered physical entity.

(2) A logical equivalent, G ("All nonblack things are nonravens"), of H is supported by conjunctive proposition $B = B_1B_2$ (where B_1 = "a is not black" and B_2 = "a is not a raven").
(3) G is entailed and thereby supported by C, "a is a white shoe."[15]

The problem arises because a plausible deductivist criterion of evidential support prescribes treating logical equivalents, here H and G, alike for support by evidence. Hence B (and, with mild further assumptions, C), in supporting G, should support H. But this is counterintuitive. Probabilistic proposals observe that B_1, "a is nonblack," is vastly more probable than A_1, "a is a raven," since there are vastly more nonblack things than ravens. Concomitantly, B will be vastly more probable than A. This pertinent observation and the resulting unsurprisingness of B make for its low relevance. To resolve the paradox formally, the known resolution proposals must also make some irrelevance assumptions. Hosiasson-Lindenbaum (1940, p. 138, ll. 5–8) assumes H to be probabilistically independent from each of A_1 and B_1. All known proposals use some familiar relevance measure or other to conclude that A and B are each positive to H, although B is positive only to a negligible degree. This explains the pretheoretical intuition that B is irrelevant to H, which extends to the irrelevance of C in the augmented story. Theorem 9 assures us that the suggestive ordinal result "B is less positive to G or H than A" holds under any possible measure satisfying POP-D.[16]

5.6 Features-in-Use of Various Relevance Measures

Measures contrast by various features that may be found desirable or undesirable, absolutely or in given uses, and normatively or descriptively. For instance, **D** must equally value prior-to-posterior shifts $0.2 \mapsto 0.4$, $0.5 \mapsto 0.7$, and $0.8 \mapsto 1$. By contrast, **L**- or **LL**-values for a target probability that increases from 0.999 to certainty, 1, must be (infinitely) higher than for an increase from 0.001 to 0.999. In this subsection, some properties affecting purely ordinal comparisons are discussed:

Extreme valuation of extreme relevance: Measures **L**, **Z**, and their ordinal equivalents attain the maximal (respectively, minimal) values of their ranges if and only if relevance is extremely positive (respectively, extremely negative). This property pleases anyone to whom conclusive argument is the strongest argument.

Amount symmetry: Measures **Q** and **QL** are amount-symmetric (section 5.1). So is the measure $R_{\mathbf{C}}(E \succ H)_P =_{df} P(EH) - P(E) \cdot P(H)$ (Carnap, 1950), which, however, violates POP-D. Among measures that are not amount-symmetric, **D** always makes the less probable of mutually relevant A and B more relevant by absolute amount to the more probable than vice versa. Formally: If not $(A \perp B)_P$, then $|R_{\mathbf{D}}(A \succ B)_P| \gtreqless |R_{\mathbf{D}}(B \succ A)_P|$ iff $P(A) \gtreqless P(B)$. **L** works likewise, except for extreme negativity.[17]

Intuitions on amount symmetry are strong only in the case of asymmetric extreme positivity. Example: Let A = "15 divides k," and B = "3 divides k," where k is only known to be a natural number ≥ 15. Evidently enough, A entails B, B does not entail A, and indeed $P(A) < P(B)$. We now have, for instance, $R_{\mathbf{D}}(A \succ B)_P > R_{\mathbf{D}}(B \succ A)_P$, in line with A's unilaterally extreme positivity to B. We also have $R_{\mathbf{Q}}(A \succ B)_P = R_{\mathbf{Q}}(B \succ A)_P$, which many will balk at, because it ignores the entailment asymmetry.

The next three features concern *conditional operation homomorphisms*, which are sometimes unconditional. Homomorphisms are structure-preserving mappings and are essential to predictability of many ordinal and all quantitative relevance properties of logical complexes. Here they map Boolean operations on propositions to equally simple, or at any rate to relatively simple, arithmetical operations on relevance values: inversion, addition, and weighted mean. Any relevance measure will afford some homomorphisms conditional on suitably narrow constraints, but most such constraints are not practically useful. Nor will all homomorphisms, conditional or otherwise, be desirable in practice. Let us look at three of them.

Relevance-functionality of negation: This refers to the inversion of relevance values r under negation: to $-r$ when $\iota_{\mathsf{T}} = 0$, to $1/r$ when $\iota_{\mathsf{T}} = 1$. It engages two distinct features: negation in target position and negation in source position. For source proposition E and target proposition H, target inversion means evidence E *always* speaks as strongly *for* hypothesis H as it speaks *against* $\neg H$. Professed opinion usually commends unconditional inversion in target position and recommends its denial for source position (cf. Eells & Fitelson, 2002). Measures **D**, **L**, and **Z** share this pair of positive and negative properties.

Conditional compositionality for conjunction: Lemma 3 (section 5.4) about **L** and **LL** exhibits the gold standard for useful conjunctive compositionality, in source position. The qualitative theorem 6 (section 3) presages its usefulness. No other familiar measures quite match this. For truly stark contrast, consider Carnap's **C**, which affords compositionality $R_{\mathbf{C}}(AB \succ H)_P = R_{\mathbf{C}}(A \succ H)_P + R_{\mathbf{C}}(B \succ H)_P$ when $P(A \vee B) = 1$. This is neither practically useful nor intuitive.

Conditional compositionality for disjunction: Here **C** is very counterintuitive: $R_C(A \vee B \succ H)_P = R_C(A \succ H)_P + R_C(B \succ H)_P$ when $P(AB) = 0$. In the case $P(AB) = 0$, intuitive, useful compositionality situates the relevance that source proposition $A \vee B$ has for some target C *between* the disjuncts' relevances. The Big Three and their isotonic transforms meet this requirement. The disjunction's relevance is always some *convex combination* (i.e., a "mixture" or weighted mean) of the disjunct relevances.

Theorem 10.

(a) If $P(AB) = 0$, then $R_T(A \vee B \succ C)_P = \alpha R_T(A \succ C)_P + (1 - \alpha)\, R_T(B \succ C)_P$, where $0 \le \alpha \le 1$ for **T** = **D**, **Q** and, if neither A nor B is extremely relevant to C, also for **T** = **L**.

(b) The mixture weight α is $\dfrac{P(A)}{P(A) + P(B)}$ for **T** = **D**, **Q**, and it is $\dfrac{P(A \mid \neg C)}{P(A \mid \neg C) + P(B \mid \neg C)}$ for **T** = **L**.

The Big Three share this mixture property with the "expected utility" concept; see Savage (1954) and chapter 8.2 by Peterson (this handbook). This commonality makes apt the characterization of each measure as a "quasi-utility," a term that Good (1983), with a different motivation, had applied to **LL**. Quasi-utilities quantify the usefulness of "source" propositions toward promoting the "target" proposition in a credal context. Under this instrumental perspective, the relation between evidence and hypothesis is a special case of a more general rhetorical relationship. Establishment of a source proposition that is positively relevant to a target proposition is a resource toward the rhetorical end of establishing the target proposition.

6. Applications in Psychology

Besides uses in philosophy of science,[18] there are psychological applications of relevance reasoning. The best known among them involve quantitative or merely ordinal judgment tasks and have a common feature. Experimental subjects appear, in some contexts, to answer requests for posterior probability or odds judgments with replies that are consistent with their being covert relevance judgments. Prominent instances of this are the phenomenon of "prior probability (or base rate) neglect" and the "conjunction fallacy."[19]

The "neglect" phenomena, when they arise, may be seen as instances of Bayes' Theorem in odds form (section 5.3). In this instance, prior odds for H are tacitly set to 1 (i.e., assume equal probability of H and $\neg H$). These odds fail to reflect information supplied on the experimenter's instruction sheet that should warrant a skewed prior, say, with $P(H) \gg P(\neg H)$, before observation E impinges. Arithmetically, such information neglect will equate posterior odds with the likelihood ratio value for relevance.

The conjunction fallacy arose most famously in the following example from the late 1970s. A fictitious person, Linda, was introduced on the experimental instruction sheet with something like the following description, E: Linda, aged 31, is single, outspoken, and bright; has majored in philosophy; and has been demonstrating for equal opportunity causes. Participants were then requested to rank the probabilities of Linda being a bank teller (B), an active feminist (F), and an active feminist bank teller (BF)—presumably in the light of information E. A surprisingly large proportion of respondents ranked BF above B, which would contravene the probability law $P(BF \mid E) \le P(B \mid E)$. Now, E was designed to be "representative" of F and "unrepresentative" of B. Many probabilists would explicate representativeness of a property Φ by the relevance of statements such as "Linda is Φ" to a reference description E, noting that F is positive to E and B negative. If one assumes that people responded to the experimenter's request with covert relevance judgments, the seemingly anomalous ranking is entirely feasible and indeed natural. See Tentori, Crupi, and Russo (2013) and Cevolani and Crupi (2015) for leads on relevance solutions of the conjunction fallacy.[20]

Finally, maximization of relevance (understood as increasing cognitive effects and reducing processing costs) has been conjectured in relevance theory to be a general principle structuring both cognition and communication. Relevance theory is popular in linguistics (Wilson & Sperber, 2004) but differs from the accounts of evidential relevance presented here. Empirically, this theory has been applied to the Wason selection task in Sperber, Cara, and Girotto (1995). Moreover, there is evidence that the measure delta-p = $P(H \mid E) - P(H \mid \neg E)$ (which is not discussed above) serves as an explication of participants' assessments of the relevance of E for H (see Skovgaard-Olsen, Singmann, & Klauer, 2016b). In Skovgaard-Olsen, Singmann, and Klauer (2016a), the relationship between this notion of relevance and indicative conditionals ("If A, then C") was investigated.

7. Conclusion

Our discussion of qualitative relevance relations has emphasized the importance of conditional independence assumptions. In practice, this means that one should seek out such independencies in one's domain of application.

Our discussion of relevance measures has shown the continued importance of conditional independence for several ordinal and compositionality properties of relevance. Another useful condition for compositionality, specifically of disjunctive evidence, was established for the major measures: the condition of mutual extreme negativity of disjuncts. More generally, our result on task-relative ordinal equivalence of the major measures provides an assurance that many ordinal relevance judgments are not unreasonably measure dependent.

If one must quantify relevance, the log-likelihood ratio **LL** is the default choice. It is the weakest sufficient statistics and has useful conditional additivity for conjunctive evidence. The relevance quotient **Q** and its logarithm **QL** are useful, for one, when relevance directionality is intrinsically unspecified. For special purposes, measures have been commended that may violate the desideratum POP-D discussed in section 5.4. Some philosophers indeed recommend measure pluralism to express diverse aspects of the relevance relation quantitatively. See Hájek and Joyce (2008) on both points.

Notes

***Editorial note*: Arthur Merin died on May 24, 2019, shortly after finishing this chapter. Final minor revisions were undertaken by the editors.**

1. See chapter 4.1 by Hájek and Staffel (this handbook). To a good extent, the considerations of this chapter could be carried out in terms of other representations of degrees of credence; see chapter 4.7 by Dubois and Prade and chapter 5.3 by Kern-Isberner, Skovgaard-Olsen, and Spohn (both in this handbook).

2. The juxtaposition AB represents the conjunction of A and B.

3. $(A \perp C \mid \pm B)_P$ can be weakened to (A not *neg* $C \mid \pm B)_P$ (W. E. Johnson; cf. Keynes, 1921). Theorem 5 follows from a result of Reichenbach (1956), which can be stated thus: If positive, neutral, and negative relevance polarities are represented by 1, 0, and –1, respectively, then, given $(A \perp C \mid \pm B)_P$, the relevance polarity of A to C is the arithmetical product of the relevance polarities of A to B and of B to C.

4. The specific benefit of conditional independence is relevance-additivity under lemma 3, section 5.4.

5. The language used in glosses is often counterfactual, particularly when referring to the "likelihood," $P(D|H)$, of H in view of D as the probability that D "would have had," given H. On likelihoods in statistics see, for example, Edwards (1972).

6. The label "Transparency" stands for "no unavoidable surprises." Hidden propositions and jumps would be surprises.

7. Thus, $R_{\mathbf{KO}}(E \succ H)_P = \tanh[\tfrac{1}{2} \log_e R_{\mathbf{LL}}(E \succ H)_P]$, that is, **KO** is just a transform of **LL**.

8. This is Carnap's 1950 name. Some authors call it "probability ratio measure" and use the label **R**, which, however, has distinct earlier referents.

9. A function $f\colon \mathbb{R} \to \mathbb{R}$ is *isotonic* iff $a < b$ implies $f(a) < f(b)$, for any $a, b \in \mathbb{R}$.

10. Reminder: $O(H) =_{\mathrm{df}} P(H)/P(\neg H)$, $P(H) = O(H)/[O(H) + 1]$, and $O(H \mid E) =_{\mathrm{df}} P(H \mid E)/P(\neg H \mid E)$.

11. Crupi and Tentori label POP-D "Final Probability," LIK-D "Law of Likelihood," MCIE "Modularity for Conditionally Independent Data," and MDHD "Disjunction of Alternative Hypotheses."

12. "Only if" parts for POP-D and LIK-D are in Steel (2003).

13. For other measures validating POP-D, among them **Z**, see Crupi and Tentori (2016). A uniqueness result for **Z** is presented in Crupi and Tentori (2013).

14. Unlike in statistics, the likelihood concept is extended by philosophers also to mutually compatible and nonexhaustive $\{H_i\}$.

15. (3) is an inessential embellishment that makes things vivid.

16. Applications of theorem 9 also include the generalization, to all measures validating POP-D, of purely ordinal results such as the following *proposition* (Merin, 1999): Let $A \dot{\vee} B =_{\mathrm{df}} (A \vee B) \wedge \neg(A \wedge B)$ represent exclusive disjunction, and assume $(A \perp B \mid \pm H)$ and $1 = \iota_{\mathsf{L}} < R_{\mathsf{L}}(A \succ H)_P,\ R_{\mathsf{L}}(B \succ H)_P < \infty$. Then $\iota_{\mathsf{L}} < R_{\mathsf{L}}(A \vee B \succ H)_P < \max[R_{\mathsf{L}}(A \succ H)_P,\ R_{\mathsf{L}}(B \succ H)_P] < R_{\mathsf{L}}(AB \succ H)_P$, and $R_{\mathsf{L}}(A \dot{\vee} B \succ H)_P < R_{\mathsf{L}}(A \vee B \succ H)_P$, while none of $R_{\mathsf{L}}(A \dot{\vee} B \succ H)_P \gtreqless \iota_{\mathsf{L}}$ are ruled out. The result shows, for one, that the defeasible "not both" implicature of assertions "A or B," which is commonly held to be inferred from the nonassertion of "A and B," can diminish the evidential relevance of inclusive "A or B" in issue-based discourse.

17. This is closely connected to what Roberto Festa calls "Matthew properties," referring to the biblical saying, "For onto every one that hath shall be given" (see Festa & Cevolani, 2016).

18. For more detail on these, see the surveys, some of them extensive, by Huber (2007), by Hájek and Joyce (2008), Crupi (2015), and Crupi and Tentori (2016).

19. See Kahneman (2011, chapters 15–16) on the two phenomena; see also chapter 1.2 by Evans and chapter 8.5 by Hertwig and Kozyreva (both in this handbook).

20. Another application of evidential relevance (see, e.g., Merin, 1999), which involves no cross-purposes in the interpretation of instructions, is to the theory of meaning for natural-language function words like "and," "or," "not," "but," "even," "also," "some," and "many."

References

Carnap, R. (1950). *Logical foundations of probability*. Chicago, IL: University of Chicago Press. 2nd ed. 1962.

Cevolani, G., & Crupi, V. (2015). Subtleties of naïve reasoning: Probability, confirmation, and verisimilitude in the Linda paradox. In M. Bianca & P. Piccari (Eds.), *Epistemology of ordinary knowledge* (pp. 211–230). Cambridge, England: Cambridge Scholars.

Crupi, V. (2015). Confirmation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/win2016/entries/confirmation/

Crupi, V., Chater, N., & Tentori, K. (2013). New axioms for probability and likelihood ratio measures. *British Journal for the Philosophy of Science*, *64*, 189–204.

Crupi, V., & Tentori, K. (2013). Confirmation as partial entailment: A representation theorem in inductive logic. *Journal of Applied Logic*, *11*, 364–372.

Crupi, V., & Tentori, K. (2016). Confirmation theory. In A. Hájek & C. Hitchcock (Eds.), *The Oxford handbook of probability and philosophy* (pp. 650–665). Oxford, England: Oxford University Press.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Edwards, A. W. F. (1972). *Likelihood*. Cambridge, England: Cambridge University Press.

Eells, E., & Fitelson, B. (2002). Symmetries and asymmetries in evidential support. *Philosophical Studies*, *107*, 129–142.

Festa, R., & Cevolani, G. (2016). Unfolding the grammar of Bayesian confirmation: Likelihood and antilikelihood principles. *Philosophy of Science*, *84*, 56–81.

Fitelson, B., & Hartmann, S. (2015). A new, Garber-style solution to the problem of old evidence. *Philosophy of Science*, *82*, 712–717.

Glymour, C. (1980). *Theory and evidence*. Princeton, NJ: Princeton University Press.

Good, I. J. (1950). *Probability and the weighing of evidence*. London, England: Griffin.

Good, I. J. (1960). Weight of evidence, corroboration, explanatory power, information, and the utility of experiments. *Proceedings of the Royal Statistical Society B*, *22*, 319–331.

Good, I. J. (1983). *Good thinking*. Minneapolis: University of Minnesota Press.

Hájek, A., & Joyce, J. M. (2008). Confirmation. In S. Psillos & M. Curd (Eds.), *The Routledge companion to philosophy of science* (pp. 115–128). Abingdon, England: Routledge.

Heckerman, D. E. (1988). An axiomatic framework for belief updates. In J. F. Lemmer & L. N. Kanal (Eds.), *Uncertainty in artificial intelligence* (Vol. 2, pp. 11–22). Amsterdam, Netherlands: North Holland.

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

Huber, F. (2007). Confirmation and induction. *Internet encyclopedia of philosophy*. Retrieved from https://iep.utm.edu/conf-ind/

Jeffrey, R. C. (2004). *Subjective probability: The real thing*. Cambridge, England: Cambridge University Press.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kemeny, J. G., & Oppenheim, P. (1952). Degree of factual support. *Philosophy of Science*, *19*, 307–324.

Keynes, J. M. (1921). *A treatise on probability*. London, England: Macmillan.

Merin, A. (1999). Information, relevance and social decision-making: Some principles and results of decision-theoretic semantics. In L. S. Moss, J. Ginzburg, & M. De Rijke (Eds.), *Logic, language, and computation* (Vol. 2, pp. 179–221). Stanford, CA: CSLI Publications.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufman.

Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.

Savage, L. J. (1954). *The foundations of statistics*. New York, NY: Wiley. Augmented ed. Dover Publications 1972.

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016a). The relevance effect and conditionals. *Cognition*, *150*, 26–36.

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016b). Relevance and reason relations. *Cognitive Science*, *41*(Suppl 5), 1202–1215.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*, 31–95.

Steel, D. (2003). A Bayesian way to make stopping rules matter. *Erkenntnis*, *58*, 213–227.

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability vs. inductive confirmation. *Journal of Experimental Psychology: General*, *142*, 235–255.

Wilson, D., & Sperber, D. (2004). Relevance theory. In L. R. Horn & G. L. Ward (Eds.), *The handbook of pragmatics* (pp. 607–632). Oxford, England: Blackwell.

The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.

The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455