Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Applied Logic ●●● (●●●●) ●●●–●●●

JOURNAL OF APPLIED LOGIC

www.elsevier.com/locate/jal

Conditionals and consequences

Henry E. Kyburg Jr. [a], Choh Man Teng [b], Gregory Wheeler [c,*]

[a] Departments of Philosophy and Computer Science, University of Rochester, Rochester, NY 14627, USA
[b] Institute for Human and Machine Cognition, University of West Florida, Pensacola, FL 32502, USA
[c] Artificial Intelligence Center – CENTRIA, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

Abstract

We examine the notion of conditionals and the role of conditionals in inductive logics and arguments. We identify three mistakes commonly made in the study of, or motivation for, non-classical logics. A nonmonotonic consequence relation based on evidential probability is formulated. With respect to this acceptance relation some rules of inference of System P are unsound, and we propose refinements that hold in our framework.

Keywords: Nonmonotonic logic; Evidential probability; System P; Sub-P logics; Imprecise probability

1. Three mistakes

> Pure Mathematics is the class of all propositions of the form '*p* implies *q*'... And logical constants are all notions definable in terms of the following: Implication, the relation of a term to a class of which it is a member ... [40, p. 3].

Thus begins the precursor of *Principia Mathematica*, Russell's *Principles of Mathematics*, and thus begins the sad and confusing twentieth century tale of implication.

1.1. Implication

Russell and Whitehead found it handy to introduce several special symbols, "$\wedge$", "$\vee$", and "$\supset$" among them. In the domain of what has come to be known as "mathematical logic", these are functions, like "$+$" and "$\times$" that combine two objects to yield a third object of the same kind. A reasonable translation of "$p \supset q$" is "if p then q", or "p only if q". This preserves, grammatically, the possibility of iteration. Just as we can grammatically refer to the sum of a and (the sum of b and c), so we can sensibly say, "if p then (if q then r)". Such locutions have a sensible place in mathematical arguments, and must be accounted for.

Like "$\vee$" and "$\wedge$", "$\supset$" has a truth functional semantics. Note that neither the English "or" nor the English "and" has a truth functional semantics, any more than "if ... then $---$" has. One shouldn't expect it to; English is a rich and flexible language.

* Corresponding author.
E-mail addresses: kyburg@cs.rochester.edu (H.E. Kyburg), cmteng@ihmc.us (Choh Man Teng), greg@di.fct.unl.pt (G. Wheeler).

1570-8683/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.jal.2006.03.014

In any case, an implication is a *relation* between two objects. So Russell is perfectly right: mathematics (and logic) concern the relation of implication, i.e., entailment, i.e., the question of what follows from what.

"Logical constants" are another matter. Unfortunately for the history of logic and the peace of mind of many philosophers, Russell and Whitehead did not pronounce their symbol "$\supset$" "only if", but "implies". This is nonsensical on a number of grounds. The English verb "implies" has an important place in mathematical talk, but it relatively rarely corresponds to the locution "if … then — — —". "Only if" or "if … then — — —" is a function: like "and" and "or" it takes objects and combines them to produce a third object. The verb "implies" is a different part of speech: it is a predicate expressing a relation between two objects, like "is greater than". It does not generate a third object. "$7 > 5$" is not odd or even, and surely not greater than 2. Similarly, while "if p then (if q then r)" not only makes sense, but has clear and natural truth conditions (p and q can't be true when r is false), "p implies (q implies r)" is unparsible.

Russell suggests that "the assertion that q is true or p false turns out to be strictly equivalent to 'p implies q' " [40, p. 15] and goes on to point out that of any two propositions, one must imply the other, that false propositions imply all propositions, and that true propositions are implied by all propositions. (These are just what have come to be known as the "paradoxes of (material) implication".) It is the truth functional constant, appearing as "$\supset$" in *Principia Mathematica*, and pronounced "implies", that has troubled logicians ever since.

In any event, the oddities noted already by Russell led others to explore the possibility of a non-truth functional connective that would play the role of Russell's "material implication", without being subject to the "paradoxes."[1] The way out seemed (fascinatingly) clear: pronounce "$p \supset q$" as "p implies q" and it is natural to consider "$p \supset q$" to be an *implication*—not in the usual sense of the word, of course, else one would ask "of what?", but in a sense peculiar to logic. One can't get rid of the old sense, though, and so to mark the distinction, we refer to "$p \supset q$" as a truth functional implication, or a *material* implication. Since "$p \supset q$" clearly doesn't exhaust the possibilities of implication [!], we can go on to consider other kinds, beginning in 1918: intuitionist implication [19], strict implication [30], relevant implication [3], linear implication [14], and so on [39].

This raises a fundamental question about the study of the relation of implication. Approaching the study of this relation through the study of a conditional connective, we contend, has yielded more confusion than insight. We trace most of this confusion to three mistakes that are commonly made when studying the relation of implication in terms of a conditional connective. The first mistake is to confuse a predicate term for a functional term.

1.2. Grammatical confusion

Some conditionals are straight-forward, and straight-forwardly truth functional. In the course of a proof we may consider alternatives, and say "If the first alternative holds, then m and n are relatively prime." From here we may go on to say that the first alternative holds, and therefore …. Or we may go on to say that m and n are not relatively prime, and therefore the first alternative cannot hold. There are few mysteries here. "If … then — — —" is acting as a statement connective, a function, and what counts are the truth values of its components.

There are also conditionals with indicative sentences as components that are (arguably) not truth functional. For example, "If the temperature reaches 50 degrees Fahrenheit, then the ice will begin to melt," expresses (as ordinarily uttered) something more than, quite different from, "Either it is false that the temperature reaches fifty degrees, or it is true that the ice will begin to melt." The conditional is being used to *express a connection* between the temperature and the melting.

Indicative conditionals—conditionals in which the components are indicative sentences—are not the focus of most work on conditionals. Most people seem to be interested in subjunctive or counterfactual "conditionals". But whether or not it is truth-functional, "if … then — — —" must at least be a sentential connective. For instance, "If it rains in Pensacola, then the ground will get wet in Pensacola" is an English sentence constructed from two sentences, *it rains in Pensacola* and *the ground will get wet in Pensacola*. Yet, overlooking this elementary condition for a conditional is another source of confusion.

Consider as an example the (presumed false) conditional,

(1) "If Oswald had not shot Kennedy, then someone else would have [shot Kennedy]."

[1] See Quine's lucid but unsympathetic discussion on pp. 28–33 of [37].

"If ... then — — —" occurring in sentence (1) is not, on its surface at any rate, a sentential connective. "Oswald had not shot Kennedy," in the sense in which it appears as the antecedent of (1), is not a sentence, nor is "Someone else would have [shot Kennedy]". It takes both of these components, occurring within the gaps of this if-then construction, to make a sentence.

In a recent book on conditionals, [8], Jonathan Bennett writes that the sentence

(2) If the British had not attacked Suez, Soviet troops would not have entered Hungary a few weeks later.

has nested within it these two expressions:

Con-1: *The British had not attacked Suez*
Con-2: *The Soviet troops would not have entered Hungary a few weeks later* [8, p. 5],

the latter of which is not even a sentence. Bennett waves away this problem, however. According to him, the subjunctive structure of (2) *paraphrases* a sentence (although not a sentence of English!), namely

(3) O_2 (the British did not attack Suez, Soviet troops did not enter Hungary a few weeks later),

where "if ... then — — —" is replaced by O_2, and the subjunctive expressions Con-1 and Con-2 are replaced by "corresponding" indicative sentences. Indeed, Bennett takes as a rough *criterion* of what he calls "the second sort" of conditionals (the "subjunctive" conditionals) that the consequent often contains the word "would". But just as dependably this very condition—containing the word "would"—precludes the possibility that the consequent is a sentence, and hence requires that the expressions occurring in the conditional be reinterpreted. There is something dubious, however, about taking (2) and interpreting it in terms of a novel two-place connective and two sentences, none of which appear in the original. For the evaluation of conditional expressions in natural language is highly variable with respect to the tense, aspect, and analytic mood of the constituent expressions.[2] Thus this move to isolate a *new* sentential connective, O_2, is an arbitrary choice masked by the substantial role played by interpretation. Nevertheless, this does allow us free reign to explore and invent the formal properties of O_2. This leads us to the third confusion over conditionals, to which we turn next.

1.3. Psychologism

Although everybody recognizes logic as a theory of what follows from what, many people *simultaneously* think of it as (part of) a theory of reasoning. *Psychologism*, the view that logical laws are "the laws of thought" [9] was deplored by Frege and rightly so:

> The ambiguity of the word 'law' is fatal here. In one sense it states what is, in the other it prescribes what should be. Only in the latter sense can the laws of logic be called laws of thought... But the expression 'law of thought' tempts us into viewing these laws as governing thinking in the same way as the laws of nature govern events in the external world. They can then be nothing other than psychological laws, since thinking is a mental process. And if logic were concerned with these psychological laws then it would be a part of psychology... [But on this view] truth is reduced to the *holding as true* of individuals. In response I can only say: *being true* is quite different from *being held as true*, whether by one, or by many, or by all, and is in no way to be reduced to it. There is no contradiction in something being true which is held by everyone as false. I understand by logical laws not psychological laws of *holding true*, but laws of *being true* [12, p. xv].

A contemporary who agrees with Frege (on this score) is Gilbert Harman. In *Change in View* [16] and most recently in "The Problem of Induction" [17], he argues at length that it is mistaken to view deductive logic as "a theory of

[2] For example, notice that the past-tense of the verb phrase in the sentence "*He stopped calling*" is not preserved in "*If he stopped calling, she would be more receptive*".

reasoning" [17, p. 3]. We agree with Harman on this point and also agree that it is a mistake to view nonmonotonic logic as a theory of inductive reasoning, even though a theory of inductive reasoning is what many nonmonotonic logicians seem to be after.

But, writes Harman,

> ... to call deductive rules 'rules of inference' is a real fallacy, not just a terminological matter. It lies behind attempts to develop relevance logics or inductive logics that are thought better at capturing ordinary reasoning than classical deductive logic does, as if deductive logic offers a partial theory of ordinary logic [17, p. 5].

Curiously, Harman claims that it is a "category mistake" to speak of inductive arguments. On this point, we disagree. We shall argue later that there is a perfectly good theory of inductive argument, just as there is a perfectly good theory of deductive argument, though we agree fully with Harman (and with Frege) that neither theory has much to do with reasoning. So long as the focus is on inductive arguments, we maintain, there is no category mistake behind the study of inductive logic.

Granted, it is not hard to understand the interest in deviant logics: as computer scientists we want to devise systems that will, like people, go beyond what is entailed by the evidence, that will focus on relevant conclusions, that will accommodate many *arguments* that do not conform to the classical deductive model, but that people regard as "good".

These interests have led to various nonmonotonic logics. In addition, they have led to attempted formalizations of non-truth functional connectives modeled to some extent on the "material implication" connective of Russell and Whitehead. It seems *almost* natural to explore the possibility of axiomatizing various new conditionals. This opens the door to an unbounded proliferation of fashionable formalisms.

However, serious confusion arises when we take these formalisms to be about reasoning. It is one thing to propose to apply some logic to model a feature of reasoning since this proposal is then subject to the same type of constraints that attend any other proposal to apply mathematics, one of which is the evaluation of how well the formal model fits the problem domain. There are several features of belief fixation and belief change that give pause to viewing logical methods as the right branch of mathematics from which to construct a formal model of rational belief fixation and rational belief change [16]. Yet the proposal may be articulated and evaluated; there is no category mistake involved in advancing such a project. Nevertheless, it is quite a different matter to view the subject of logic itself to be reasoning, since psychologism in effect denies that there is a need to judge the fit between the formal framework and the problem domain. The "inference-implication" fallacy applies to psychologism, not to applied logic in general. While it is a mistake to regard the study of inductive logics as the study of human inference, it is likewise mistaken to regard the study of inductive logics to be predicated on this very error.

2. Logic is not linguistics

Despite the fact that many contemporary treatments of conditionals rest on one or more of these mistakes, there is obviously an important topic buried here. People use the English locution "if ... then $---$" and its relatives very frequently, of course, but, more than this, those locutions occur in important contexts. These are contexts in which what follows from what is at issue, where the symbol "$\vdash$" might be more appropriate than "$\supset$", but also, and more importantly, in which the issue concerns the weight of the evidence. Harman is no doubt right to protest the psychologizing of sentential connectives, but that does not mean that inductive logic rests on a "category mistake", or that the treatment of the implicational or evidential import of conditionals is misguided. A remark then on applied logic.

Part of the bargain of using a formal language to represent a more complicated target domain (such as natural language) is that we must settle on the key features of the domain to represent and leave the others aside. Presumably, our choice will be guided by a question we have about *that* domain: in answering questions about the movement of the planets, we are not thrown by representing them as point-masses. Nevertheless, this arrangement becomes more complicated when the target domain is a more expressive language, since we must interpret the expressions in the target language solely by the restricted terms provided by our formal language. The nagging problem for a *logical* analysis (as opposed to a *linguistic* analysis) of conditionals is to decide on the appropriate set of semantic features on which to base an account. The extensive literature on conditionals suggests that there are no precise features for indicative conditionals—or, conversely, that no class of expressions corresponding to a fixed set of logical features is co-extensive with the class of indicative conditional sentences in English. A challenge for friends of conditionals then is to explain why we shouldn't think this the end of the matter.

Nevertheless, there are particular *uses* of conditional expressions in English that are of logical interest, namely when we are discussing what follows from what. And there is an important epistemic relation that, we will argue, is a candidate for logical analysis. It is instructive to see how this topic is obscured by confusing a psycho-linguistic analysis of conditional sentences for a logical analysis of an epistemic function that is sometimes expressed in the use of natural language conditional sentences.

Suppose we restrict ourselves to *indicative conditionals*, which are "if ... then − − −" sentences whose components are expressions in the indicative mood. Anthony Gillies [13] has argued that there "is a serious *epistemological* problem" raised by interpreting indicative conditionals as truth functional, as is proposed by [20,31]. To motivate the problem, Gillies asks us to consider two detectives discussing the alibis of 3 murder suspects, all of whom work at a mansion, two of whom are members of grounds staff and the remaining a member of the house staff. Upon ruling out the house staff, Detective A errs by asserting

(4) "Therefore: If a member of the grounds staff did it, then it was the driver" [13, p. 589].

whereat his partner B replies,

(5) "It's not so that if a member of the grounds staff did it, then it was the driver. After all, it might still be the gardener who did it" [13, p. 589].

Gillies thinks that the problem behind interpreting "if ... then − − −" in (4) and (5) truth functionally is that it saddles us with epistemic commitments that aren't reasonable to hold. The "denial in [(5)]", Gillies writes,

> would commit you to accepting that a member of the ground staff did it and it was definitely not the driver. This is too strong a commitment, and would render your reply unwarranted. But given your information, [(5)] is not only the reasonable thing to *say*, it is the rational thing to *believe*—and believing that a member of the grounds staff did it and it was definitely not the driver would be decidedly irrational [13, p. 590].

But what precisely is being denied in (5)? Given the setup for the example, what makes (5) reasonable to assert is that $\not\models_{PL} (g \vee d) \supset d$. Detective B is (baroquely) pointing out that his partner forgot about the gardener.

The role of interpretation here is relevant, since Gillies is claiming that the existence of a counter-model to $\models_{PL} (g \vee d) \supset d$—namely when $(g \vee d) \wedge \neg d$ is satisfied—must express an epistemic commitment if we interpret "if ... then − − −" as truth functional. But the worry over epistemic commitments is a red herring. The word "therefore" in (4) already signals that what is at issue is implication, and the denial expressed by (5) is a denial that (4) is valid. That $(g \vee d) \wedge \neg d$ is satisfiable is the reason that $\not\models_{PL} (g \vee d) \supset d$, full stop. This is simply a concise way to represent the constraints of the state of the case and the reason that detective A is mistaken to assert (4). This is the gist of what the English sentences in (4) and (5) are being used to express. How these English sentences effect this behavior is a matter for linguistics and psychology to settle, not logic.

If we confuse linguistic analysis for logical analysis, then interpreting the "$\supset$" within mathematical proofs itself becomes suspect. A counter-example expressed by writing out the denial of some conditional $p \supset q$ does not entail an epistemic commitment of author or reader to the content of p. That the author or reader's epistemic commitments may change after writing or understanding a proof is one thing. That an assertion of the proof itself represents such a commitment is quite another. The study of how natural language performs in this task is interesting and important, but it is not the domain of logic.

The interesting topic buried here is entailment. Our exhortation to study notions of consequence rather than conditionals *per se* is grounded in the observation that conditional expressions in natural language typically involve considerably more structure than the relation of implication. Even when an "if ... then − − −" expression is used to identify the relation of implication, such as in the example above, there may be linguistic features that have nothing to do with illustrating the underlying relation of implication that the sentences are being used to express.

Moreover, even if we restrict ourselves to subsets of English that *do* admit translation to sentences of first-order logic, only an impoverished fragment of English is decidable [36]. This is not to deny that one can isolate important fragments of conditional logics that map nicely to various notions of consequence, as in [7]. Rather, we're arguing that "natural language friendly" logic is of very limited value: inventing logical connectives to fit data from the semantics or pragmatics of natural language "if ... then — — —" expressions simply does not advance our understanding of entailment.

Logic is the study of what follows from what. Just as there is a perfectly good, objective and useful classical logical/mathematical theory about what follows from what, so, we would maintain, there is the potential of an objective and useful logical/mathematical theory of what provides good evidence for what. This observation naturally takes us to probability, which we consider next.

3. Conditional probability

One kind of "conditional" that clearly makes sense is conditional probability. Almost all authors regard almost all probabilities to be conditional probabilities, even though many people follow Kolmogorov in *introducing* probability as a one-place function. However, $P(B|A) = P(BA)/P(A)$ is *not* a definition of conditional probability, since $P(B|A)$ is undefined when $P(A) = 0$. Rather, this ratio is an axiom concerning the primitive expression $P(B|A)$, and not as clear a one as $P(BA) = P(A)P(B|A)$.

There are many interpretations of probability, and what we are to make of conditional probability depends on the interpretation we adopt. For present purposes, however, the only *kind* of interpretation that we need be concerned with is one that assigns probabilities to pairs of sentences. (We will suggest later that probabilities should be assigned to sentences and *sets* of sentences, but for the moment the simpler idea will suffice.) Thus $P(h|e)$ is the probability assigned to h conditional on e.

Many have viewed probability as a means to capture the evidential relationship between sentences. One reason for this view is that conditional probability appears suited to represent the defeasible character of evidential relations: unlike implication, good evidential support for a claim may well be undone by additional evidence. Whereas implication is (weakly, positively) monotonic, evidential support is not.

Still, it should be noted that the probability function $P(\cdot|\cdot)$ itself is monotone: if $A \subseteq B$ then $P(A|\cdot) \leqslant P(B|\cdot)$; If A is a smaller set of possibilities than B, then the probability of A must be less than (or equal to) B. Of course the second position of $P(\cdot|\cdot)$ may be thought of as non-monotone, in some sense, since if $A \not\equiv B$ then $P(\cdot|A)$ may be greater than, less than, or equal to $P(\cdot|A \cap B)$; Conditioning on a smaller set of possibilities may either increase or decrease the conditional probability. It is natural then to ask whether there is a connection between nonmonotonic logics—which feature nonmonotonic consequence relations—and conditional probability.

The path for studying nonmonotonic conditionals opened after the independent and distinct work of [34] and [28], where each provided a probabilistic semantics for satisfying the axioms of System P, a system first discussed in [21]. System P consists of a number of axioms and rules of inference that are taken to be a conservative core any nonmonotonic system should contain. Let $\mid\!\sim$ be a nonmonotonic consequence relation: $\alpha \mid\!\sim \beta$ denotes that together with the suppressed background knowledge, α is good, but not necessarily certain, evidence for β. Let $\vee$, $\wedge$, $\rightarrow$ and $\leftrightarrow$ be standard connectives in a classical propositional logic, $\rightarrow$ being the truth functional conditional. Let $\models \alpha$ denote α is valid. $\models \ulcorner \alpha \rightarrow \beta \urcorner$ can be equivalently expressed as $\alpha \models \beta$.[3] The axiomatization of system P is as follows.

$\alpha \mid\!\sim \alpha$ [**Reflexivity**]

$$\frac{\models \ulcorner \alpha \leftrightarrow \beta \urcorner;\ \alpha \mid\!\sim \gamma}{\beta \mid\!\sim \gamma} \qquad [\textbf{Left Logical Equivalence}]$$

$$\frac{\models \ulcorner \alpha \rightarrow \beta \urcorner;\ \gamma \mid\!\sim \alpha}{\gamma \mid\!\sim \beta} \qquad [\textbf{Right Weakening}]$$

$$\frac{\alpha \mid\!\sim \beta;\ \alpha \mid\!\sim \gamma}{\alpha \mid\!\sim \ulcorner \beta \wedge \gamma \urcorner} \qquad [\textbf{And}]$$

[3] We follow Quine [37] in using quasi-quotation (corners) to specify the *forms* of expressions in our formal language. Thus $\ulcorner S \leftrightarrow T \urcorner$ becomes a specific biconditional expression on the replacement of S and T by specific formulas of the language.

$$\frac{\alpha \mathrel{|\!\sim} \gamma;\ \beta \mathrel{|\!\sim} \gamma}{\ulcorner \alpha \vee \beta \urcorner \mathrel{|\!\sim} \gamma} \qquad [\textbf{Or}]$$

$$\frac{\alpha \mathrel{|\!\sim} \beta;\ \alpha \mathrel{|\!\sim} \gamma}{\ulcorner \alpha \wedge \beta \urcorner \mathrel{|\!\sim} \gamma} \qquad [\textbf{Cautious Monotonicity}]$$

Although Pearl bases his approach on infinitesimal probabilities and Lehmann and Magidor base theirs on a non-standard probability calculus, each shares the view that "acceptance" or "full belief" is to be identified with maximal probability, a view that has been developed by [4,11], defended in [15] and studied by [6], where the latter includes an important limitative result.

However, there is little reason to think that the notion captured in these models bears any resemblance to the evidential relation that figures in inductive arguments. The lottery paradox [22] casts doubt on the plausibility of the axiom [**And**], for example [18,42].[4]

Thus, we deny that System P is the conservative core of nonmonotonic logic. Nevertheless, we do think that the tools of logic and probability are appropriate to study evidential relations. On our conception of nonmonotonic logic, what is to be studied are relations that hold in arguments between the accepted evidence and a nonmonotonically derived conclusion. Next we will discuss a nonmonotonic consequence relation based on evidential probability, and examine the axioms and rules of inference of System P in light of this interpretation.

4. Acceptance based on evidential certainties

We are interested in the logical structure of the set of conclusions that may be obtained by nonmonotonic or inductive inference from a body of evidence. Epistemic states can be modeled as acceptance of sentences. These are sentences that ought to be believed with respect to an acceptable chance of error. For reasons discussed elsewhere [24], we must make a sharp distinction between the set of sentences constituting the *evidence*, and the set of nonmonotonically or inductively *inferred* sentences. The evidence itself may be uncertain, but we shall suppose that it carries less risk of error than the knowledge we derive from it. Let us take the set of sentences constituting the *evidence* to be Γ_δ and the set of sentences constituting our accepted knowledge Γ_ϵ, where $\delta \leqslant \epsilon$; δ represents the risk of error (degree of corrigibility) of our evidence, and ϵ represents the risk of error of what we infer from the evidence. For example, we obtain the error distribution of a method of measurement from a statistical analysis of its calibration results; but we take that distribution for granted, as *evidence*, when we regard the result of a measurement as giving the probable value of some quantity.

4.1. Evidential probability

The idea behind evidential probability [22,26] is that probabilities should reflect known relative frequencies, to the extent that we know them, and also be tailored to the particular case at hand. Probabilities are thus attributed to sentences relative to a *set of sentences* Γ_δ representing our background knowledge. The value of a probability, in view of the limitations of our knowledge, is an *interval* rather than a real number. In this framework, all probabilities are conditional, in the sense that the probability of a statement is given conditional on a finite set of sentences taken as the background knowledge or evidence.

We will follow the treatment of probability in [26]. Let $\mathcal{L}$ be the language; the domain of the probability function is $\mathcal{L} \times \wp(\mathcal{L})$, and its range is intervals $[l, u]$. The language allows the expression of statistical knowledge, in the form $\ulcorner \%\bar{\eta}(\tau, \rho, l, u) \urcorner$, where $\bar{\eta}$ is a sequence of logical variables, and the statement as a whole says that among the models of Γ_δ satisfying ρ, between a fraction l and a fraction u also satisfy τ. (The formulas τ and ρ are not arbitrary formulas, but formulas belonging to canonical lists of target formulas and reference formulas, respectively.) It is on sets of statements that include statistical knowledge that probabilities are based.

We construe probability as a metalinguistic relation (on analogy with provability) and thus provide a *definition* for probability rather than axioms for it. We represent the probability of the statement S, given the background knowledge Γ_δ, by $\mathrm{Prob}(S, \Gamma_\delta) = [l, u]$. The basic ingredients for a potential probability statement concerning S, given background

[4] It should be noted the "full belief" model is susceptible to the transfinite versions of the lottery paradox [32].

knowledge Γ_δ, are a triple of terms, α, τ, ρ, where $\ulcorner S \leftrightarrow \tau(\alpha)\urcorner$, $\ulcorner \rho(\alpha)\urcorner$, and $\ulcorner \%x(\tau(x), \rho(x), l, u)\urcorner$ are all in the set of evidence statements Γ_δ.

Given a statement S, there are many statements of the form $\ulcorner \tau(\alpha)\urcorner$ known to have the same truth value as S, where α is known to belong to some reference class ρ and where some statistical connection between ρ and τ is known. In the presence of $\ulcorner S \leftrightarrow \tau(\alpha)\urcorner$ and $\ulcorner \rho(\alpha)\urcorner$ in Γ_δ, the statistical statement $\ulcorner \%\bar{\eta}(\tau, \rho, l, u)\urcorner$ is a candidate for determining the probability of S. The problem is that there are many such triples α, τ, ρ. This is the classic problem of the reference class.

In [26] we offered three principles for eliminating excess candidates.

We say that two statistical statements $\%x(\tau(x), \rho(x), l, u)$ and $\%y(\tau'(y), \rho'(y), l', u')$ *differ* just in case neither $[l, u] \subseteq [l', u']$ nor $[l', u'] \subseteq [l, u]$ holds.

1. If two statistical statements differ and the first is based on a marginal distribution, while the second is based on the full joint distribution, ignore the first. This gives conditional probabilities pride of place *when they conflict with the equivalent unconditional probabilities*.
2. If two statistical statements differ and the second employs a reference formula that logically entails the reference formula employed by the first, ignore the first. This embodies the well-known principle of *specificity*.

Those statistical statements we are not licensed to ignore we will call *relevant*. A smallest set of statistical statements that contains every relevant statistical statement that *differs* from a statement in it will be said to be *closed under difference*.

3. The probability of S is the shortest cover of any non-empty set of relevant statistical statements closed under difference; alternatively it is the intersection of all such covers. This embodies the (controversial) principle of *strength*.

An example may help. Suppose the intervals mentioned in the set of relevant statements are [0.20,0.30], [0.25,0.35], [0.22,0.37], [0.40,0.45], [0.20,0.80], [0.10,0.90], [0.10,0.70]. There are three sets that are closed under difference; {[0.20,0.30], [0.25,0.35], [0.22,0.37], [0.40,0.45]}, {[0.20,0.80], [0.10,0.70]}, and {[0.10,0.90]}. The first set contains an interval that is included in another interval, and the third set is a singleton. The probability is [0.20,0.45].

It can be shown that evidential probability is sound: if $\mathrm{Prob}(\alpha, \Gamma_\delta) = [l, u]$ then the proportion of models of Γ_δ in which α is true lies between l and u.

It is natural to ask about the relationship between evidential probability and the more usual axiomatic approaches. Since evidential probability is interval valued, it can hardly satisfy the usual axioms for probability. However, it can be shown that if Γ_δ is consistent, then there exists a one place classical probability function P whose domain is $\mathcal{L}$, and that satisfies $P(S) \in \mathrm{Prob}(S, \Gamma_\delta)$ for every sentence S of $\mathcal{L}$.

The story with regard to conditional probability is a little more complicated. Of course all probabilities are essentially conditional, and that is clearly the case for evidential probability. However, as Levi showed in 1977 [29], there may be no one place classical probability function P such that $P(S|T) = P(S \wedge T)/P(T) \in \mathrm{Prob}(S, \Gamma_\delta \cup \{T\})$. That is, the evidential probability of S relative to the corpus Γ_δ supplemented by evidence T need *not* be compatible with any conditional probability based on Γ_δ alone. This is a consequence, as Levi observes, of our third principle, the principle of strength.

While one should not abandon conditioning lightly, it should be observed that in many instances there is no conflict, and in particular evidential probability can adjudicate between classical inference and Bayesian inference when there is conflict. For more details see [25].

4.2. ϵ-Acceptability

We would like our body of accepted knowledge, or practical certainties, to be objectively justified by the evidence. Given a body of evidence Γ_δ, every statement S of our language has a unique probability and this probability is based on frequencies that are known in Γ_δ to hold in the world. In the limiting case where nothing is known about a statement, its probability is the whole probability interval $[0, 1]$. Note, however, that there are a priori statistical statements, such as "$\%x$(x reflects relative frequency of A in parent population of B's within $k/(2\sqrt{n})$, x is a sample of n B's, $1-1/k^2$, 1)", which is true whatever the frequency of A in B. It is natural to suggest that it is worth accepting a statement S as known if there is only a negligible chance that it is wrong. Put in terms of probability, we might say that it is reasonable to accept a statement into Γ_ϵ when the maximum probability of its denial relative to what we take as evidence, Γ_δ, is less than or equal to a fixed "small" value ϵ. This reflects—and is in part motivated by—the theory of testing statistical hypotheses. We thus have the following definition of Γ_ϵ, our body of accepted knowledge or practical certainties, in terms of our body of evidence Γ_δ:

$$\Gamma_\epsilon = \big\{S : \exists\, l, u\big(\mathrm{Prob}(\ulcorner \neg S \urcorner, \Gamma_\delta) = [l, u] \wedge u \leqslant \epsilon\big)\big\},$$

or alternatively,

$$\Gamma_\epsilon = \big\{S : \exists\, l, u\big(\mathrm{Prob}(S, \Gamma_\delta) = [l, u] \wedge l \geqslant 1 - \epsilon\big)\big\}.$$

Given Γ_δ, we say a sentence S is *ϵ-accepted* if $S \in \Gamma_\epsilon$.

Note that the small number "ϵ" is to be construed as a fixed number, rather than a variable that approaches a limit. Both Ernest Adams [1,2] and Judea Pearl [34,35] have sought to make a connection between high probability and logic, but both have taken probabilities *arbitrarily* close to one as corresponding to knowledge. These approaches involve matters that go well beyond what we may reasonably suppose to be available to us as empirical enquirers. In real life we do not have access to probabilities arbitrarily close to one. Thus we have chosen to follow the model of hypothesis testing in statistics: we reject a hypothesis (accept its complement) when the chance of error in doing so is less than a fixed finite amount (ϵ), relative to a body of evidence that suffers a chance of no more than δ of being in error.

5. Rules of inference for acceptance

We will examine whether the axioms and rules of System P conform to the notion of ϵ-acceptability, and in those cases where they do not, we will investigate in the next section alternative versions that do hold.

Let us identify the consequence relation $\mid\!\sim$ with ϵ-acceptability, that is, $\alpha \mid\!\sim \beta$ is interpreted as β is ϵ-accepted when α is added to the body of evidence Γ_δ; in other words, the probability $\mathrm{Prob}(\beta, \Gamma_\delta \cup \{\alpha\})$ has a lower bound of at least $1-\epsilon$. The background knowledge Γ_δ is taken to be omitted by convention in the above rules of inference, analogous to the omission in the specification of conditional probability, where $P(B|A)$ denotes the probability of B given A *and* the assumed but not mentioned background knowledge.

With the above interpretation of $\mid\!\sim$ and the soundness of evidential probability, we can show that the axiom schema [**Reflexivity**] is valid and the two rules of inference [**Left Logical Equivalence**] and [**Right Weakening**] are sound.

1. [**Reflexivity**] α is given a probability of [1, 1] when added to Γ_δ and therefore is ϵ-acceptable for any ϵ.
2. [**Left Logical Equivalence**] Two equivalent formulas α and β are true in the same models, and therefore the probability of γ with respect to β is the same as the probability of γ with respect to α.
3. [**Right Weakening**] Given that the truth functional conditional $\ulcorner \alpha \rightarrow \beta \urcorner$ is valid, β obtains whenever α obtains. Thus, the probability of β with respect to γ is at least as much as the probability of α with respect to γ.

The remaining three rules [**And**], [**Or**] and [**Cautious Monotonicity**] are not sound rules of inference in our framework, as we will show below.

5.1. [**And**]

The lower bound of the probability of the conjunction $\ulcorner \beta \wedge \gamma \urcorner$ is not higher than and can be lower than the smaller of the two lower bounds of the probabilities of β and of γ. For instance, the probability that a jelly bean (α) is red (β) is at least 90%, and the probability that a jelly bean is round (γ) is at least 90%. The lower bound of the probability that a jelly bean is both red *and* round ($\ulcorner \beta \wedge \gamma \urcorner$) is at most (and may be lower than) 90%. In the case that redness and roundness of jelly beans are independent, the lower bound of this probability is 81%.

The following joint probability distribution illustrates this example. Each line of the table denotes the probability of the combination of statements α, β and γ (or their negations).

α	β	γ	Probability	α	β	γ	Probability
0	0	0	0	1	0	0	1/100
0	0	1	0	1	0	1	9/100
0	1	0	0	1	1	0	9/100
0	1	1	0	1	1	1	81/100

Given a lower bound requirement of 90%, we have $\alpha \mid\!\sim \beta$ (90%) and $\alpha \mid\!\sim \gamma$ (90%) but not $\alpha \mid\!\sim \ulcorner \beta \wedge \gamma \urcorner$ (81%).

5.2. [**Or**]

A counter-example demonstrates the failure of this rule. Suppose the probability that a red jelly bean (α) is apple-flavored (γ) is at least 90%, and the probability that a round jelly bean (β) is apple-flavored is at least 90%. Actually all apple-flavored jelly beans are red *and* round. In this case the lower bound of the probability that a jelly bean is apple-flavored given that it is red *or* round ($\ulcorner \alpha \vee \beta \urcorner$) is 81.8%.

A joint probability distribution for this counter-example is as follows.

α	β	γ	Probability	α	β	γ	Probability
0	0	0	0	1	0	0	1/11
0	0	1	0	1	0	1	0
0	1	0	1/11	1	1	0	0
0	1	1	0	1	1	1	9/11

Given a lower bound requirement of 90%, we have $\alpha \mid\!\sim \gamma$ $(\frac{9/11}{10/11})$ and $\beta \mid\!\sim \gamma$ $(\frac{9/11}{10/11})$ but not $\ulcorner \alpha \vee \beta \urcorner \mid\!\sim \gamma$ $(\frac{9/11}{11/11})$.

5.3. [**Cautious Monotonicity**]

Another counter-example suffices. Again suppose the probability that a jelly bean (α) is red (β) is at least 90%, and the probability that a jelly bean is round (γ) is at least 90%. All jelly beans that are not red are round. In this case the lower probability that a red jelly bean ($\ulcorner \alpha \wedge \beta \urcorner$) is round is 88.9%.

The following joint probability distribution illustrates this scenario.

α	β	γ	Probability	α	β	γ	Probability
0	0	0	0	1	0	0	0
0	0	1	0	1	0	1	1/10
0	1	0	0	1	1	0	1/10
0	1	1	0	1	1	1	8/10

Given a lower bound requirement of 90%, we have $\alpha \mid\!\sim \beta$ $(\frac{9/10}{10/10})$ and $\alpha \mid\!\sim \gamma$ $(\frac{9/10}{10/10})$ but not $\ulcorner \alpha \wedge \beta \urcorner \mid\!\sim \gamma$ $(\frac{8/10}{9/10})$.

6. Refined rules of inference

We may ask under what conditions the above three unsound rules of system P do hold. One way to weaken the rules is to modify the premises of the rules so that the rules only apply in a restricted setting.

Note that in each of the unsound rules, two nonmonotonic consequence relations are involved in the premises, whereas in the sound rules only one such relation is involved. Secondly note that the rules of inference would be sound if we replace all nonmonotonic relations ($\mid\!\sim$) by their classical deductive counterparts ($\models$), either in both the premises and conclusions or only in the premises of each rule.

We may modify the unsound rules by changing one of the nonmonotonic relations in the premises to a classical deductive relation, for instance

$$\frac{\alpha \models \beta;\ \alpha \mid\!\sim \gamma}{\alpha \mid\!\sim \ulcorner \beta \wedge \gamma \urcorner} \qquad [\textbf{And*}]$$

$$\frac{\alpha \models \gamma;\ \beta \mid\!\sim \gamma}{\ulcorner \alpha \vee \beta \urcorner \mid\!\sim \gamma} \qquad [\textbf{Or*}]$$

$$\frac{\alpha \models \beta; \alpha \mid\!\sim \gamma}{\ulcorner \alpha \wedge \beta \urcorner \mid\!\sim \gamma} \qquad \textbf{[Cautious Monotonicity*]}$$

The above modified rules of inference are sound in our framework. For the rules [**And***] and [**Or***], the modification is symmetrical, that is, the change from $\mid\!\sim$ to $\models$ may be applied to either of the two premises. For [**Cautious Monotonicity**] the change from $\mid\!\sim$ to $\models$ may also be applied to either of the two premises, each change resulting in a distinct and sound rule of inference.

Below we will discuss in more detail these refinements and their relationship to the original rules of inference in System P. Note that the restrictions we impose here on the unsound rules of inference are only some of the ways the rules may be made sound in accordance with the notion of ϵ-acceptability. We may consider other ways of placing restrictions on the rules. Some of these variations will be discussed below.

6.1. [**And**] *rule and adjunction*

The failure of the [**And**] rule with respect to ϵ-acceptability means that in general we do not have adjunction in our framework. This may strike some people as disastrous. After all, adjunction is a basic form of inference. In mitigation, we point out that the modified rule [**And***] shows that we *do* have adjunction under certain restricted circumstances, when one of the statements to be adjoined follows strictly from the premises.

To the extent that we are thinking of an argument as *supporting* its conclusion, any argument requires simultaneous (adjunctive) acceptance of its premises. Any doubt that infects the conjunction of the premises rightly throws a shadow on the conclusion. Except for notational convenience, any argument may be taken to have a single premise.

The persuasiveness of the argument depends on the empirical support given to the conjunction of the premises. Too many premises, each only *just* acceptable, will not provide good support for a conclusion even when the conclusion does validly follow from the premises. Furthermore, although the conclusion should not be accepted, there may be no particular premise that could be singled out for rejection.

Modal logics have often been used to characterize the logical structure of nonmonotonic inference. Classical modal systems, with their weaker non-normal constraints on the models, do not necessarily satisfy adjunction [10,33,41]. In particular, classical modal systems without the axiom schema

(**C**) $(\Box\phi \wedge \Box\psi) \rightarrow \Box(\phi \wedge \psi)$

may be used to characterize nonmonotonic inference structures similar to that discussed here [5,27]. For a more complete discussion of adjunction and its role in logic, see [23].

6.2. [**Or**] *rule*

For [**Or***], the lower bound of the probability of γ relative to $\ulcorner \alpha \vee \beta \urcorner$ is at least as high as the lower bound of the probability of γ with respect to β alone.

Proof of [Or*]. Let $\alpha \models \gamma$ and let the probability of γ with respect to β be at least l. We need to show that the probability of γ with respect to $\ulcorner \alpha \vee \beta \urcorner$ is also at least l.

Let $\mathcal{M}(\alpha)$ denote the number of models (of Γ_δ) satisfying α. Again Γ_δ will be omitted for convenience. We will also omit quasi quotes in the following. From our premises we have

1. $\frac{\mathcal{M}(\gamma \wedge \alpha)}{\mathcal{M}(\alpha)} \in [1, 1]$, and therefore $\mathcal{M}(\gamma \wedge \alpha) = \mathcal{M}(\alpha)$; and
2. $\frac{\mathcal{M}(\gamma \wedge \beta)}{\mathcal{M}(\beta)} \in [l, u]$ for some l and u.

Now the proportion of models satisfying γ among those satisfying $\ulcorner \alpha \vee \beta \urcorner$ is

$$\begin{aligned}\frac{\mathcal{M}(\gamma \wedge (\alpha \vee \beta))}{\mathcal{M}(\alpha \vee \beta)} &= \frac{\mathcal{M}(\gamma \wedge \alpha) + \mathcal{M}(\gamma \wedge \beta) - \mathcal{M}((\gamma \wedge \alpha) \wedge (\gamma \wedge \beta))}{\mathcal{M}(\alpha \vee \beta)} \\ &= \frac{\mathcal{M}(\alpha) + \mathcal{M}(\gamma \wedge \beta) - \mathcal{M}(\gamma \wedge \alpha \wedge \beta)}{\mathcal{M}(\alpha \vee \beta)}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathcal{M}(\alpha) + \mathcal{M}(\gamma \wedge \beta) - \mathcal{M}(\alpha \wedge \beta)}{\mathcal{M}(\alpha) + \mathcal{M}(\beta) - \mathcal{M}(\alpha \wedge \beta)} \\
&= \frac{\mathcal{M}(\gamma \wedge \beta) + k}{\mathcal{M}(\beta) + k} \\
&\geqslant \frac{\mathcal{M}(\gamma \wedge \beta)}{\mathcal{M}(\beta)} \\
&\geqslant l.
\end{aligned}
$$

Thus, the probability of γ with respect to $\ulcorner \alpha \vee \beta \urcorner$ is at least l. □

For example, suppose we now know that not just 90% but *all* red jelly beans (α) are apple-flavored (γ). We also know that the probability that round jelly beans (β) are apple-flavored is at least 90%. Since adding red jelly beans (which are all apple-flavored) to a pile of round jelly beans (which are 90% apple-flavored) is not going to decrease the overall probability of apple-flavored jelly beans in the pile (compared to the original pile of round jelly beans), the lower probability that a jelly bean that is red *or* round is apple-flavored is at least as high as the lower probability that a round jelly bean is apple-flavored. In the worst case all apple-flavored jelly beans are red *and* round, as in the counter-example for [**Or**]. With the additional constraint imposed by [**Or***] that all red jelly beans are apple-flavored, the lower probability that a jelly bean that is red *or* round is apple-flavored in this case is 90%.

Note that the original [**Or**] rule holds in the special case where α and β are mutually exclusive, that is

$$\frac{\models \ulcorner \neg\alpha \vee \neg\beta \urcorner; \alpha \mathrel{|\!\sim} \gamma; \beta \mathrel{|\!\sim} \gamma}{\ulcorner \alpha \vee \beta \urcorner \mathrel{|\!\sim} \gamma} \qquad [\textbf{X-Or}]$$

We denote this rule by [**X-Or**]. Both [**X-Or**] and [**Or***] are sound. The difference between the two rules corresponds to the difference between exclusive disjunction and boolean disjunction, respectively.

6.3. [**Cautious Monotonicity**]

In [**Cautious Monotonicity***] the modification is not symmetrical. In the form shown at the start of Section 6, γ retains the same probability relative to $\ulcorner \alpha \wedge \beta \urcorner$ as relative to α alone. Alternatively, the rule may be modified as follows.

$$\frac{\alpha \mathrel{|\!\sim} \beta; \alpha \models \gamma}{\ulcorner \alpha \wedge \beta \urcorner \mathrel{|\!\sim} \gamma} \qquad [\textbf{Cautious Monotonicity**}]$$

In this case the probability of γ relative to $\ulcorner \alpha \wedge \beta \urcorner$ is [1, 1].

In addition the "inverse" of [**Cautious Monotonicity**]

$$\frac{\alpha \mathrel{|\!\sim} \beta; \ulcorner \alpha \wedge \beta \urcorner \mathrel{|\!\sim} \gamma}{\alpha \mathrel{|\!\sim} \gamma} \qquad [\textbf{Cumulative Transitivity}]$$

is also unsound in our framework. This rule is closely related to but weaker than [**Cautious Monotonicity**] in the sense that it holds in some systems in which [**Cautious Monotonicity**] does not hold, for instance Reiter's default logic [38]. Again we can obtain a sound inference rule from this unsound rule by replacing one of the $\mathrel{|\!\sim}$'s in the premises by $\models$.

7. Conclusion

In considering the notion of conditionals and the role conditionals play in logic and arguments, we have raised several methodological issues at the heart of applied logic. We've catalogued three mistakes that are commonly made in the study of, or motivation for, non-classical logics and then cited a motivation offered for the formal study of *epistemic conditionals* as an illustrative example. We do not think that the application of non-classical logics to model epistemic notions must involve any one of these mistakes, however. We think that epistemic relations are open to mathematical study and we think that logical methods and tools are appropriate. In this light we formulated a non-monotonic consequence relation based on evidential probability. With respect to this acceptance relation we observe that some rules of inference of System P are unsound, specifically [**And**], [**Or**] and [**Cautious Monotonicity**]. We proposed refined versions of these rules that hold in our framework and discuss their connections to the original rules and other variations of the rules.[5]

References

[1] E. Adams, Probability and the logic of conditionals, in: J. Hintikka, P. Suppes (Eds.), Aspects of Inductive Logic, North-Holland, Amsterdam, 1966, pp. 265–316.
[2] E. Adams, The Logic of Conditionals, Reidel, Dordrecht, 1975.
[3] A.R. Anderson, N. Belnap, Entailment, vol. 1, Princeton University Press, Princeton, NJ, 1975.
[4] H. Arló-Costa, Bayesian epistemology and epistemic conditionals, Journal of Philosophy 98 (11) (2001) 555–598.
[5] H. Arló-Costa, First order extensions of classical systems of modal logic; the role of the Barcan schemas, Studia Logica 71 (1) (2002) 87–118.
[6] H. Arló-Costa, R. Parikh, Conditional probability and defeasible inference, Journal of Philosophical Logic (2005).
[7] H. Arló-Costa, S. Shapiro, Maps between conditional logic and non-monotonic logic, in: B. Nebel, C. Rich, W. Swartout (Eds.), Principles of Knowledge Representation and Reasoning, Morgan Kaufmann, San Mateo, CA, 1992, pp. 553–565.
[8] J. Bennett, Philosophical Guide to Conditionals, Oxford University Press, Oxford, 2003.
[9] G. Boole, The Mathematical Analysis of Logic, Oxford University Press, Oxford, 1948. Originally published by Cambridge University, 1847.
[10] B. Chellas, Modal Logic, Cambridge University Press, Cambridge, 1980.
[11] B.C. Van Fraassen, Fine-grained opinion, probability, and the logic of belief, Journal of Philosophical Logic 95 (1995) 349–377.
[12] G. Frege, The Foundations of Arithmetic, Basil Blackwell, Oxford, 1893, j. l. austin (trans). edition, (1950).
[13] A. Gillies, Epistemic conditionals and conditional epistemics, Nous 38 (4) (2004) 585–616.
[14] J.-Y. Girard, Linear logic, Theoretical Computer Science 50 (1987) 1–101.
[15] J.Y. Halpern, Reasoning about Uncertainty, MIT Press, Cambridge, MA, 2003.
[16] G. Harman, Change in View, MIT Press, Cambridge, MA, 1986.
[17] G. Harman, S.R. Kulkarni, The problem of induction, in: Philosophy and Phenomenological Research, in press.
[18] J. Hawthorne, Nonmonotonic conditionals that behave like conditional probabilities above a threshold, Journal of Applied Logic, this issue.
[19] A. Heyting, Intuitionism, an Introduction, vol. 3, North-Holland Co., Amsterdam, 1955 (1971).
[20] F. Jackson, On assertion and indicative conditionals, The Philosophical Review 80 (1979) 565–589.
[21] S. Kraus, D. Lehman, M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, Artificial Intelligence 44 (1990) 167–207.
[22] H.E. Kyburg Jr., Probability and the Logic of Rational Belief, Wesleyan University Press, Middletown, CT, 1961.
[23] H.E. Kyburg Jr., The rule of adjunction and reasonable inference, Journal of Philosophy 94 (3) (1997) 109–125.
[24] H.E. Kyburg Jr., Real logic is nonmonotonic, Minds and Machines 11 (2001) 577–595.
[25] H.E. Kyburg Jr., Bayesian inference with evidential probability, in: W. Harper, G. Wheeler (Eds.), Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr., King's College, London, 2006.
[26] H.E. Kyburg Jr., Choh Man Teng, Uncertain Inference, Cambridge University Press, Cambridge, 2001.
[27] H.E. Kyburg Jr., Choh Man Teng, The Logic of Risky Knowledge, Electronic Notes in Theoretical Computer Science, vol. 67, Elsevier Science, Amsterdam, 2002.
[28] D. Lehman, M. Magidor, What does a conditional knowledge base entail? Artificial Intelligence 55 (1990) 1–60.
[29] I. Levi, Direct inference, Journal of Philosophy 74 (1977) 5–29.
[30] C.I. Lewis, A Survey of Symbolic Logic, University of California Press, Berkeley, CA, 1918.
[31] D. Lewis, Probabilities of conditionals and conditional probabilities, The Philosophical Review 85 (3) (1976) 297–315.
[32] P. Maher, Acceptance without belief, in: PSA 1990, vol. 1, 1990, pp. 381–392.
[33] R. Montague, Universial grammer, Theoria 36 (1970) 373–398.
[34] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Francisco, CA, 1988.
[35] J. Pearl, System Z: A natural ordering of defaults with tractable applications to default reasoning, in: Theoretical Aspects of Reasoning about Knowledge, 1990, pp. 121–135.
[36] I. Pratt-Hartmann, Fragments of language, Logic, Language and Information 13 (2) (2004) 207–223.
[37] W.V.O. Quine, Mathematical Logic, second ed., Harvard University Press, Cambridge, MA, 1940.
[38] R. Reiter, A logic for default reasoning, Artificial Intelligence 13 (1980) 81–132.
[39] G. Restall, An Introduction to Substructural Logics, Routledge, London, 2000.
[40] B. Russell, Principles of Mathematics, Cambridge University Press, Cambridge, 1903.
[41] D. Scott, Advice in modal logic, in: K. Lambert (Ed.), Philosophical Problems in Logic, Reidel, Dordrecht, 1970, pp. 143–173.
[42] G. Wheeler, Rational acceptance and conjunctive/disjunctive absorption, Journal of Logic, Language and Information 15 (2006), in press.

[5] This material is based upon work supported by NSF award IIS-0328849, NASA NNA04CK88A, ONR N00014-03-1-0516, and FCT award SFRH/BPD-13699-2003. The authors would like to thank Horacio Arló-Costa, Jim Hawthorne, Jon Kvanvig, David Makinson, Eric Pacuit, Teddy Seidenfeld and two anonymous referees for comments on an earlier draft of this paper.