Nonmonotonic Conditionals that Behave Like Conditional Probabilities Above a Threshold

James Hawthorne

University of Oklahoma
Department of Philosophy
605 Dale Hall Tower
Norman, OK 73019, USA
hawthorne@ou.edu

Keywords. nonmonotonic logic, nonmonotonic conditionals, consequence relations, conditional probabilities, probabilistic thresholds

Abstract. I'll describe a range of systems for nonmonotonic conditionals that behave like conditional probabilities above a threshold. The rules that govern each system are probabilistically sound in that each rule holds when the conditionals are interpreted as conditional probabilities above a threshold level specific to that system. The well-known *preferential* and *rational consequence relations* turn out to be special cases in which the threshold level is 1. I'll describe systems that employ weaker rules appropriate to thresholds lower than 1, and compare them to these two standard systems.

1 Introduction

I will describe a range of nonmonotonic conditionals that behave like conditional probabilities above a threshold. More precisely, let r be a fixed real number greater than 1/2, and let P be any conditional probability function defined on a language for predicate logic with identity. Consider the conditional |~ defined as follows: B|~A holds just in case P[A | B] ≥ r. Let's call |~, as just defined, the *r-level consequence relation associated with conditional probability function* P. More generally, the *r-level consequence relations* are just those associated with at least one conditional probability function P at threshold level r. I will characterize r-level consequence relations for various values of r in terms of logical rules – rules like, "if (B·C)|~A and (B·¬C)|~A, then B|~A", which are only about the conditional expressions, and say nothing of probabilities. It turns out that the logical rules that these conditionals satisfy are mostly weaker versions of the logical rules for the two best-known logics of nonmonotonic conditionals – i.e., the logics of the *preferential consequence relations*, **P**, and of the *rational consequence relations*, which I'll call **R**.[1]

The conditionals I'll be investigating are of the kind that nonmonotonic logicians call 'consequence relations' by analogy to the *logical consequence relation*. They are

---

[1] See [4] for a thorough treatment of **P**, and [5] for the scoop on **R**. The best known semantics for these conditionals is in terms of *preferential models*, which is fully explicated in these two papers, and also in [7].

metalinguistic relations between sentences. Conditional probability functions applied to sentences are also usually taken to be metalinguistic – i.e. they are not generally taken to be part of the object language. So, the corresponding conditionals, |~, are also metalinguistic. Thus, I will call all of the conditionals under investigation here 'consequence relations'.

This paper is aimed at two audiences. It's pitched at probabilistic logicians, who may find it of interest for the way it articulates the qualitative structure of *conditional-probability-above-a-threshold*. In that regard this logic is somewhat like the logic of *Qualitative Probability* (a.k.a. *Comparative Probability*). But whereas the basic concept in that logic is the 'A *is-no-more-probable-than* B' relation, A ≥ B, the basic concept we'll be looking at here is the '*given* B, *it-is-very-probable-that* A' relation B|~A. More accurately, we will investigate a whole array of such consequence relations ranging from '*it-is-more-probable-than-not-that*' (corresponding to a threshold just over 1/2), through consequence relations for various higher threshold levels, up to '*it-is-almost-certain-that*' (corresponding to a threshold of 1). Each quantitative probability function embodies all of these qualitative notions at once. Perhaps something can be learned from disentangling them.

The other audience I'm pitching consists of logicians interested the nonmonotonic conditionals known as *consequence relations*. The systems under study have a common core, a system I'll call **O**, whose rules are weaker analogues of the rules for the well-known system **P** of *Preferential Consequence Relations*. Various ways of supplementing the rules of **O** give rise to various systems of consequence relations, including the system **P** itself and the system for the *Rational Consequence Relations*, **R**. What ties these systems together is the way in which they are embodied by (i.e. modeled in) the conditional probability functions. Indeed, every conditional probability function embodies a complete array of nonmonotonic consequence relations drawn from the systems we'll be looking at.

Here is a brief outline of how I'll proceed. First I'll specify the logic of conditional probabilities that will serve as a standard against which we'll gauge the nonmonotonic conditional logics. Next I'll set down rules for a system of consequence relations that I call **O**. Each rule for consequence relations in **O** is r-level sound for each r > 0. That is, choose any specific threshold r > 0: then by replacing each conditional expression of form B|~A in the rules of **O** by a conditional probability sentence P[A | B] ≥ r, each rule turns out to hold for every probability function P. If we add an additional simple rule to **O**, the rule known as AND, we get the preferential consequence relations **P** – even though three of the rules of **O** are much weaker than their usual counterparts associated with **P**. Add to the rules of **P** the rule known as *Rational Monotony* (RM) and we get the rational consequence relations **R**. Interestingly, it turns out that the consequence relations of system **R** are just the class of threshold-level-1 probabilistic consequence relations. That is, for each relation |~ that satisfies the rules of **R** there is a conditional probability function P such that B|~A holds just in case P[A | B] = 1; and the probability 1 part of each conditional probability function is just a consequence relation |~ that satisfies the rules of **R**.

The fact that the 1-level consequence relations are so tightly connected with the well-known rational consequence relations suggests that it may be illuminating to take a look at r-level consequence relations for values of r less than 1. We will see what additional rules, added to **O**, are probabilistically sound for various levels of the threshold r below 1. Although neither AND nor *Rational Monotony* (RM), nor the rule known as *Cautious Monotony* (CM) are probabilistically sound for thresholds r below 1, the weaker monotonicity rule called *Negation Rationality* (NR) does turn out to be probabilistically sound *for each possible threshold level* $r > 0$. The system for consequence relations gotten by supplementing **O** with *Negation Rationality* constitutes the system I'll call **Q**.

None of the rules of **Q** itself are specific to a given threshold level $r > 0$. That is, each rule of **Q** applies to all r-level consequence relations, for any given value of r. However, there are two additional rules that are closely tied to specific threshold levels. One applies whenever the threshold r is greater than some rational number $(n-1)/n$, for fixed $n \geq 2$. The other applies whenever the threshold r is no greater than the rational number $n/(n+1)$, for fixed $n \geq 2$. I call the system of consequence relations that satisfy both of these rules, for a specific value of n, **Q**(n). They behave like conditional probabilities above some threshold r such that $(n-1)/n < r \leq n/(n+1)$. These level-specific rules turn out to have a close connection to the Preface and the Lottery Paradoxes.

2 The Logic of Conditional Probabilities

When applied to propositions or sentences, probability is usually specified as a one-place function, and conditional probability is then defined in terms of this function: $P[A \mid B] = P[A{\cdot}B]/P[B]$ for $P[B] > 0$, and $P[A \mid B]$ is undefined for $P[B] = 0$. However, there is a very natural way of axiomatizing probability that takes conditional probability as primitive. It turns out that this treatment of conditional probability is closely related to logics of nonmonotonic conditionals.

Think of conditional probabilities as extending logical entailment to a conception of probabilistic truth-transmission from premise to conclusion. There are many such extensions – many such probability functions. For, unlike deductive logical entailment, the notion of probabilistic entailment may depend on what the sentences mean. Formally, the degree to which premise B probabilistically entails conclusion A relative to an *interpretation* α of a language L is represented by a function on pairs of sentences, $P_\alpha[A \mid B]$. We may think of each conditional probability function P_α as associated with some way of assigning meanings to the terms of the language, and as supplementing that with a measure on possible worlds. Thus, '$P_\alpha[A \mid B] = r$' may be taken to say that among the worlds in which sentence B is true, A is true in proportion r of them (according to some measure on worlds associated with P_α).

Alternatively, we may think of 'α' as representing a possible agent, and think of each possible agent as having (either implicitly or explicitly) some degree-of-belief function that expresses how strongly premise sentences support conclusion sentences. For an agent α, let P_α represent her conditional degree-of-belief function, given the meaning of the sentences of her language.

Whatever way one conceptualizes the conditional probability functions, the axioms for these functions specify constraints that they must respect given the meanings of logical terms (not, and , or, etc.). Here is a fairly standard set of axioms:

Definition 1: CP. Let L be a language for predicate logic with identity. Let '$\models$' be the standard logical entailment relation. A **Conditional Probability Function** (**CP Function**) on L is any function P from pairs of sentences to real numbers between 0 and 1 that satisfies the following rules.

0. There are sentences D and E such that, $P[D \mid E] < 1$;
   for all sentences A, B, C:
1. If $B \models A$, then $P[A \mid B] = 1$;
2. If $\models (B \equiv C)$, then $P[A \mid B] = P[A \mid C]$;
3. If $C \models \neg(B \cdot A)$, then either $P[(A \vee B) \mid C] = P[A \mid C] + P[B \mid C]$ or $P[D \mid C] = 1$ for every sentence D;
4. $P[(A \cdot B) \mid C] = P[A \mid (B \cdot C)] \cdot P[B \mid C]$.

Holding any sentence C fixed, each function $P[... \mid C]$ behaves just like a classical unconditional probability function as usually defined on sentences of a formal language. Furthermore, whenever $P[Y \mid C] > 0$, $P[X \mid Y \cdot C] = P[X \cdot Y \mid C] / P[Y \mid C]$, in agreement with the classical definition of conditional probability. However, the **CP** functions extend classical probability in that they remain defined even when probabilities are *conditionalized* on sentences having probability 0 – i.e. $P[A \mid (B \cdot C)]$ remains defined even when $P[B \mid C] = 0$.[2]

Perhaps a comment on the formal language I'm using is in order here before proceeding. All of the logical systems I'll describe in this paper are defined on a standard formal language L for predicate logic with identity. Nothing I'll say really hangs on this. The language could just as well have been weaker – say, that of sentential logic. But then the reader might have been left wondering whether the results only hold for the weaker language.[3]

---

[2] The **CP** functions are basically just the Popper-Field functions; however the usual axiomatization for the Popper-Field functions is more elegant in that it does not employ (or in any way presuppose) the deductive notion of logical truth or logical consequence. See [2] and [3] for details.

[3] It is also common to define probability on propositions instead of on sentences of a formal language, where a proposition is taken to be a set of possible worlds. In that case one would have to broach the issue of whether the probability functions are countably additive or only finitely additive. That issue doesn't arise here because the object language L doesn't have an expression for infinite disjunction (which would correspond to countable unions). It does, however, have existential quantifiers, which behave somewhat like infinite disjunctions. Indeed, one could add a weak kind of countable additivity axiom to the axioms of **CP**, as follows: for each open formula Fx, $P[\exists xFx \mid B] = \lim_n P[Fc_1 \vee ... \vee Fc_n \mid B]$, where the individual constants $c_1, ..., c_n, ...$, exhaust the countably infinite list of L's individual constants. However, in the context of predicate logic this axiom seems overly strong, since it effectively assumes that every individual gets named. If we don't assume that all individuals are named, the strongest claim we should want is that $P[\exists xFx \mid B] \geq \lim_n P[Fc_1 \vee ... \vee Fc_n \mid B]$. However, this already follows from the axioms of **CP**, because $B \cdot (Fc_1 \vee ... \vee Fc_n) \models \exists xFx$, so $P[\exists xFx \mid B \cdot (Fc_1 \vee ... \vee Fc_n)] = 1$, so $P[\exists xFx \mid B] \geq P[(Fc_1 \vee ... \vee Fc_n) \mid B \cdot \exists xFx] \cdot P[\exists xFx \mid B] = P[(Fc_1 \vee ... \vee Fc_n) \cdot \exists xFx \mid B] = P[\exists xFx \mid B \cdot (Fc_1 \vee ... \vee Fc_n)] \cdot P[(Fc_1 \vee ... \vee Fc_n) \mid B] = P[(Fc_1 \vee ... \vee Fc_n) \mid B]$.

3 Systems O and P

There are two complimentary ways of describing a logic for nonmonotonic consequence relations. Sometimes logicians identify such a logic in terms of its inference rules. The issue then is, "what are reasonable rules for a nonmonotonic consequence relation to follow, and what inferences can we make from some conditional claims to others based on these rules." Other times we think of such a logic as a class of possible consequence relations, all of those that satisfy certain constraints. In this mode the "rules" just express the constraints. In standard deductive logic the former mode is most usually associated with the "syntactic" proof theory of the logic, and the later mode is associated with the semantics. But labeling these two modes as "syntactic" and "semantic" is not very helpful in the logics we will be looking at. Rather, in these logics it may be best to think of the consequence relations themselves as semantic, as part of the metalanguage, just as probability functions are semantic – i.e., they are not part of the object language. That's how presentations of the well-known nonmonotonic logics **P** and **R** often treat consequence relations (e.g. see [4] and [5]).

To see the point, think about the logic of conditional probability – the system **CP** just described. The axioms of **CP** can be used to derive some conditional probability claims from others – a very useful thing when we have only partial information about a probability function (or about a class of such functions). On the other hand, we also associate with **CP** the class of all conditional probability functions – all functions on pairs of sentences that satisfy the axioms. When we think of the logic this way, the axioms play the role of constrains that must be satisfied if a function is to be considered a **CP-function**. Thus the "logic of **CP**" both tells us which functions are "in **CP**" in terms of rules that specify constraints on all such functions, and it gives us rules for deriving some probability statements from others, where the soundness of such derivations depend on the fact that all **CP**-functions are defined in terms of those very rules.

Both of these ways of looking at **CP** is "semantic" in the sense that the probability functions in **CP** are semantic functions – not part of the object language. The object language on which they are defined is generally a formal language for sentential or predicate logic. The probability functions play the role of metalinguistic, semantic predicates, in much the way that *truth-under-interpretation* is metalinguistic and semantic. So the "derivation rules" are really semantic rules that specify precisely what semantic claims (involving probabilities) can be derived from others – much as the semantic rules governing truth-value assignments can be used to derive claims about what truth-values of sentences follow from the truth-values of other sentences.

Each of the systems for nonmonotonic consequence relations I'll discuss has this same dual aspect. For example, I am about to specify "the logic" of the system I'll call **O**. (Think of **O**, represented by the letter 'O', as *system-zero*, the weakest system I'll talk about.) I will specify **O** in terms of certain semantic rules that any consequence relation must satisfy to be an **O**-relation.

Consider the set of all ordered pairs of sentences from a given language. Take any subset of it – let's call it 'α'. Any such α is a rudimentary consequence relation. We usually write these ordered pairs like this: 'B|~A', rather than like this: '<B,A>'. So when the pair <B,A> is in α, we say instead that the conditional expression 'B|~A' is in α. (One might here employ the subscripting convention we used with probability functions: just as there are various possible probability functions P_α , there are various possible consequence relations $|\sim_\alpha$.)

Many such consequence relations will be of no interest at all. They violate even the most obvious constraints on how a consequence relation should behave. The system **O** specifies some very weak semantic rules that we'll suppose any set of pairs of sentences should satisfy if it is to reasonably count as a consequence relation. The *consequence relations in* **O** are just those that satisfy the following semantic constraints.

Definition 2: O. Let L be a language for predicate logic with identity. Let '|=' be the standard logical entailment relation. An **O Consequence Relation** on L is any set of pairs of sentences that satisfies the following rules:

0. There are sentences D and E such that it's not the case that E|~D	(Non-Triviality)
1. A|~A	(REFLEX: Reflexivity)
2. if C|~B and B|=A, then C|~A	(RW: Right Weakening)
3. if B|=C and C|=B and B|~A, then C|~A	(LLE: Left Logical Equivalence)
4. if (C·B)|~A and (C·¬B)|~A, then C|~A	(WOR: Weak Or)
5. if C|~(B·A), then (C·B)|~A	(VCM: Very Cautious Monotony)
6. (C·¬B)|~B , C|~A then C|~(B·A)	(WAND: Weak And)

Some of the rules of **O** should be familiar. The non-triviality condition is not usually given. Clearly it is only violated by that one monstrous consequence relation that holds between all pairs of sentence. Reflexivity, Right Weakening, and Left Logical Equivalence are plausible conditions, satisfied by all well-known families of consequence relations. The three remaining rules are weakened versions of well-known rules for nonmonotonic consequence relations. We'll look at them in more detail in a moment.

Notice that each of the rules 0-6 is probabilistically sound at level r, for any level r you might choose. That is, fix a threshold value r. Now replace each expression of form X|~Y in these rules with the corresponding expression P[Y | X] ≥ r. Then each such rule is a theorem of probability theory – i.e., each follows from the rules of **CP**. This is obvious for rules 0-3. Rule 4 (WOR) becomes the following theorem of **CP**:

if P[A | (C·B)] ≥ r and P[A | (C·¬B)] ≥ r, then P[A | C] ≥ r.[4]

The soundness of rule 5 (VCM) is obvious, since the following is a theorem of **CP**:

---

[4] Because **CP** is a slightly non-standard axiomatization of probability, and the reader may not be familiar with it, I'll be very careful here. First observe that for **CP**, whenever Z |= (X≡Y), P[X | Z] = P[Y | Z]. (That's easy to show: suppose Z |= (X≡Y); then Z |= ¬(X·¬Y) and Z |= (X∨¬Y); so either P[D | Z] = 1 for all D (and we're done), or by rule 3, 1 = P[X∨¬Y | Z] = P[X | Z] + P[¬Y | Z] = P[X | Z] + 1 − P[Y | Z]; done.) Now suppose P[A | (C·B)] ≥ r and P[A | (C·¬B)] ≥ r. Then either P[A | C] = 1 ≥ r (done), or (from rules 3, 4, and the previous result) 1 > P[A | C] = P[(A·B)∨(A·¬B) | C] = P[A·B | C] + P[A·¬B | C] = P[A | B·C]·P[B|C] + P[A | ¬B·C]·P[¬B | C] ≥ r.

if $P[(B \cdot A) \mid C] \geq r$, then $P[A \mid (C \cdot B)] \geq r$.[5]

Rule 6 (WAND) is also sound, since:

if $P[\neg B \mid (C \cdot B)] \geq r$ and $P[A \mid C] \geq r$, then $P[(A \cdot B) \mid C] \geq r$.[6]

Thus we have the following theorem.

Theorem 1: Probabilistic Soundness of O. Choose any threshold level $r > 0$. The rules of **O** are probabilistically sound at level r.

However **O** is not probabilistically complete There are consequence relations satisfying all of these rules that cannot be represented by any conditional probability function and threshold. This follows from the fact that there are additional probabilistically sound rules not derivable from the rules of **O**. We will see some of these additional rules presently.

How do the rules of **O** compare with those of the well-known *preferential consequence relations*, characterized by the set of rules **P**? **P** has rules 1-3 (REFLEX, RW, and LLE). But **P** contains stronger versions of each of the rule 4-6. Here is a typical definition of **P**:

Definition 3: P. Let L be a language for predicate logic with identity. Let '|=' be the standard logical entailment relation. A **P Consequence Relation** on L is any set of pairs of sentences that satisfies the following rules:

0. There are sentences D and E such that it isn't the case that E|~D	(Non-Triviality)
1. A|~A	(REFLEX: Reflexivity)
2. if C|~B and B|=A, then C|~A	(RW: Right Weakening)
3. if B|=C and C|=B and B|~A, then C|~A	(LLE: Left Logical Equivalence)
4**P**. if B|~A and C|~A, then (B∨C)|~A	(OR)
5**P**. if C|~B and C|~A, then (C·B)|~A	(CM: Cautious Monotonicity)
6**P**. if C|~B and C|~A, then C|~(B·A)	(AND)

Rule 0 is usually left out of **P**. It merely eliminates the consequence relation for which each sentence is a consequence of every sentence. Now let's compare the rules of **O** to the corresponding rules of **P**.

Rules 0-3 are the same. So consider rule 4**P** (OR). This rule is probabilistically sound only when the threshold level $r = 1$.[7] OR cannot be derived from WOR plus the other **O** rules, because whereas the **O** rules are probabilistically sound at thresholds below 1, OR is not. Indeed, from rules 1-3 it is easy to show that WOR is equivalent to the following rule, which makes the relationship between OR and WOR transparent:

---

[5] Suppose $P[B \cdot A \mid C] \geq r$. Then $r \leq P[A \cdot B \mid C] = P[A \mid C \cdot B] \cdot P[B \mid C] \leq P[A \mid C \cdot B]$.

[6] Suppose for $r > 0$, $P[A \mid C] \geq r$ and $P[B \mid C \cdot \neg B] \geq r$. If $P[B \cdot A \mid C] = 1$, we're done; so suppose $P[B \cdot A \mid C] < 1$. Since $P[\neg B \mid \neg B \cdot C] = 1$ (rule 1), we have $P[B \mid \neg B \cdot C] + P[\neg B \mid \neg B \cdot C] \geq r + 1 > 1$, so $P[B \mid \neg B \cdot C] = 1$ (rule 3), so $1 = P[\neg(B \cdot \neg B) \mid C] = 1 - P[B \cdot \neg B \mid C] = 1 - P[B \mid \neg B \cdot C] \cdot P[\neg B \mid C] = 1 - P[\neg B \mid C] = P[B \mid C]$ (rules 1, 3, 4, 3). Then, $r \leq P[A \mid C] = P[(A \cdot B) \vee (A \cdot \neg B) \mid C] = P[A \cdot B \mid C] + P[A \cdot \neg B \mid C] = P[A \mid B \cdot C] \cdot P[B \mid C] + P[A \mid \neg B \cdot C] \cdot P[\neg B \mid C] = P[A \mid B \cdot C]$.

[7] See the Appendix.

4*. if |= ¬(B·C) and B|~A and C|~A, then (B∨C)|~A (XOR: exclusive or).[8]
This makes it clear that WOR is derivable from OR (together with the other **O** rules), and that OR is a strengthening of WOR.

Rule 5**P** (CM), is also probabilistically sound only when the threshold level r = 1.[9] Clearly rule 5 (VCM) can be derived from CM (and the other **O** rules). But since CM is sound only for r = 1, it cannot be derive from VCM together with the other **O** rules. The way in which CM is a strengthening of VCM is obvious.

AND, which is the **P** counterpart of **O** rule 6 (WAND) is also probabilistically sound only when the threshold level r = 1.[10] Furthermore, WAND can be derived from AND plus the other **O** rules; but since AND is sound only for r = 1, it cannot be derive from WAND plus the other **O** rules.

The relationship between WAND and AND may not seem quite obvious. To see it more clearly, notice that the condition '(C·¬B)|~B' in the antecedent of WAND is a strengthening of the condition 'C|~B' in AND.[11] It expresses the idea that "C makes B certain" – i.e., C supports B so strongly that adding any other sentence D to C cannot undermine its support for B. (We'll establish this in a moment.) AND strengthens WAND by weakening its antecedent condition '(C·¬B)|~B' to the condition 'C|~B'.

To see more clearly what an expression of form '(C·¬B)|~B' means in the context of **O**, consider the following theorem:

Theorem 2: Some Theorems of O. Let |~ be any consequence relation in **O**.

(1) The following three conditions are equivalent:
(i) (C·¬B)|~B; (ii) for all E, (C·¬B)|~E; (iii) for all D, (C·D)|~B.

(2) If (C·¬B)|~B, then, for all A, C|~A if and only if (C·B)|~A.

proof: Clause 1: Clearly, (ii) implies (i), and (iii) implies (i). To see that (i) implies (ii): suppose (i); then (C·¬B)·¬B|~B (LLE) and C·¬B|~¬B (REFLEX, RW); so C·¬B|~B·¬B (WAND); so C·¬B|~E for any E (RW). To see that (i) implies (iii): suppose (i); then for all E, C·¬B|~E (just proved); so for any D, C·¬B|~(D·B); then for any D, (C·D)·¬B|~B (VCM, then LLE) and (C·D)·B|~B (by REFLEX, RW); thus, for any D, C·D|~B (WOR).
Clause 2 of the theorem follows in one direction directly from rules WAND and VCM, and in the other direction from clause 1 (ii) and WOR.

Thus, when '(C·¬B)|~B' holds, '(C·¬B)' acts like a contradiction – i.e., '(C·¬B)' implies everything, and 'C' itself monotonically implies 'B'.

Thus, **O** is weaker than **P** precisely in that its versions of rules 4-6 (WOR, VCM, WAND) are weaker versions of the corresponding **P** rules (OR, CM, AND), versions that are satisfied at each threshold r > 0 by every conditional probability function.

One might wonder whether we need to strengthen each of the weaker **O** rules in

---

[8] Suppose WOR, and suppose |= ¬(B·C) and B|~A and C|~A. Then (by LLE) (B∨C)·¬C|~A and (B∨C)·C|~A, so B∨C|~A (WOR). Conversely, suppose XOR, and suppose C·B|~A and C·¬B|~A. Then (C·B)∨(C·¬B)|~A (XOR), so C|~A (LLE).
[9] See the Appendix.
[10] See the Appendix.
[11] In **O**, C·¬B|~B implies C|~B: for, C·B|~B (REFLEX, RW), so by WOR C|~B.

order to get the system **P**? The answer turns out to be, no! It's not hard to show that 4**P** (OR) and 5**P** (VCM) are derivable from the weaker rules (1-5) together 6**P** (AND). Thus, the *Preferential Consequence Relations* are just those consequence relations in **O** that also satisfy AND.

Theorem 3: Alternative Rules for P.
Given rules 1-3 (REFLEX, RW, LLE), rules 4**P**-6**P** (OR, CM, AND) imply rules 4-6 (WOR, VCM, WAND). Given rules 1-3, rules 4 and 5 (WOR, VCM) together with 6**P** (AND) imply rules 4**P** (OR) and 5**P** (CM). Thus, |~ is a relation in **O** that satisfies 6**P** (AND) *if and only if* |~ is a consequence relation in **P**.

proof: Given 1-3, getting 4-6 from 4**P**-6**P** is easy. So lets go the other way. Suppose 1-5 and 6**P**.
Here is how to get 4**P** (OR): Suppose B|~A and C|~A. Then (B∨C)·B|~A (LLE), so (B∨C)·B|~A∨¬B (RW); and (B∨C)·¬B|~A∨¬B (REFLEX, RW); thus B∨C|~A∨¬B (WOR). Fairly similarly, (B∨C)·C|~A (LLE), so (B∨C)·C|~A∨B (RW); also (B∨C)·¬C|~A∨B (REFLEX, RW); thus B∨C|~A∨B (WOR). From the "thus" parts of the previous two sentences, B∨C|~(A∨B)·(A∨¬B) (AND); so B∨C|~A (RW).
5**P** (CM) is easy: Suppose C|~B and C|~A. Then C|~B·A (AND), so C·B|~A (VCM).

4 Systems Q and R

The well-known *Rational Consequence Relations* are usually obtained by adding the following rule to those in **P**:

Definition 4: R. An **R Consequence Relation** on L is any **P** consequence relation that satisfies the following rule:
7**R**. if C|~A, then C|~ ¬B or (C·B)|~A (RM: Rational Monotony).

Like 4**P**-6**P** (OR, CM, AND), rule RM is probabilistically sound only for threshold level r = 1.[12]

The *Rational Consequence Relations*, **R**, are usually obtained by adding RM to the **P** rules. But in light of the previous result the usual rules of **R** are equivalent to the weaker rules 1-5 for **O** together with 6**P** (AND) and 7**R** (RM). In other words, the *Rational Consequence Relations* are just those consequence relations in **O** that satisfy AND together with RM.

It turns out that the rules in **O** ∪ (AND, RM) are not only probabilistically sound at threshold level 1. They are also probabilistically complete at level 1. That is, for each Rational Consequence Relation |~ in **R**, there is a corresponding conditional probability function P in **CP** such that 'B|~A' holds just in case P[A | B] = 1. In effect, the *Rational Consequence Relations* are just the probability 1 parts of *Conditional Probability Functions*. Or, to put it another way, given any *Rational Consequence Relation* |~, it can always be extended to a conditional probability function P by assigning P[A | B] = 1 when 'B|~A' holds and by assigning some

[12] See the Appendix.

appropriate non-negative number below 1 to P[C | D] whenever 'D|~C' fails to hold. Let's state all of this formally.

Theorem 4: Probabilistic Soundness and Completeness of R. For each **CP** function P, if |~ is the level-1 consequence relation corresponding to P (i.e. if |~ is defined by 'B|~A' holds just in case P[A | B] = 1), then |~ is in **R**. Furthermore, for each consequence relation **|~** in **R**, there is a probability function P in **CP** such that P[A | B] = 1 just in case B|~A is in **R**.

proof: Soundness is easy. Completeness takes hard work (see [1], [3], and [8]).

The system I call **Q** is the weaker analog of **R**, much as **O** is the weaker analog of **P**. Here is the definition of **Q**:

Definition 5: Q. A **Q Consequence Relation** on L is any **O** consequence relation that satisfies the following rule:
7. if C|~A then (C·B)|~A or (C·¬B)|~A (NR: Negation Rationality).

It is easy to check that NR is probabilistically sound for each threshold level.[13] Thus, all of the **Q** rules are probabilistically sound at every threshold level. NR (rule 7) is clearly a weaker analog of RM (rule7**R**), and is derivable in **R**. Indeed, NR may be derived using only rule RM together with VCM (rule 5 of **O**) together with the following rule (which is derivable in **P**):

8. if B|~A and B|~¬A, then B|~D for every sentence D (XM: excluded middle).[14]

Notice that XM is itself probabilistically sound for each threshold level greater than 1/2.[15] This rule is implied by one of a spectrum of rules that correspond to lower bounds on threshold levels. We'll now investigate systems that build on **Q** by drawing on such threshold-specific rules.

5 The Q(n) Systems

The rules of **Q** place no constraints on the value of the threshold level r > 0 required for conditional probability functions to satisfy them. That is, choose any probability function P from **CP** and any threshold level r you want. You may even choose r to be much smaller than 1/2 – even extremely close to 0. The consequence relation

---

[13] Suppose P[A | C] ≥ r. Then *either* for every D, P[D | C] = 1, so 1 = P[A·B | C] = P[A | C·B] · P[B | C], so P[A | C·B] = 1 ≥ r, *or else* r ≤ P[A | C] = P[(A·B)∨(A·¬B) | C] = P[A | C·B] · P[B | C] + P[A | C·¬B] · P[¬B | C], which cannot be if both r > P[A | C·B] and r > P[A | C·¬B].

[14] To see that NR follows from VCM, RM, and XM, suppose C|~A. (1) If C|~¬B and C|~B, then by XM, C|~B·A, so C·B|~A by VCM. (2) If C |/~ ¬B, then C·B|~A by RM. (3) If C |/~ B, then C|/~¬¬B, so C·¬B|~A by RM.

[15] Suppose r > 1/2, and suppose P[A | B] ≥ r and P[¬A | B] ≥ r. By **CP** rule 3, since P[A | B] + P[¬A | B] = 2r > 1, we have P[D | B] = 1 for all D.

corresponding to P[|] ≥ r will, nevertheless, satisfy all of the **Q** rules. We now supplement **Q** with rules that characterize various levels of probabilistic support above 1/2. For each integer $n \geq 2$ we specify a distinct "n-level logic", defined in terms of two rules that are jointly probabilistically sound for all and only threshold levels r in the range $(n-1)/n < r \leq n/(n+1)$, for $n \geq 2$.

Definition 6: Q(n). For specific integer $n \geq 2$, a **Q**(n) **Consequence Relation** on L is any **Q** consequence relation that satisfies the following two rules:

8{n}. if $(B\cdot(A_1\cdot\ldots\cdot A_n))|\sim \neg(A_1\cdot\ldots\cdot A_n)$, $B|\sim A_1$, ..., $B|\sim A_n$,
then for all D, $B|\sim D$ (PL(n): Preface Logic n);

9{n+1}. if $(B\cdot(A_1\cdot A_2))|\sim \neg(A_1\cdot A_2)$, ..., $(B\cdot(A_i\cdot A_j))|\sim \neg(A_i\cdot A_j)$, ..., $(B\cdot(A_n\cdot A_{n+1}))|\sim \neg(A_n\cdot A_{n+1})$,
then $B|\sim \neg A_1$ or ... or $B|\sim \neg A_{n+1}$ (LL(n+1): Lottery Logic n+1).

Notice that PL(n) doesn't presuppose that the A_i are distinct sentences. Thus, the rule PL(n) implies each of the rules PL(m) for $m \leq n$. Also notice that the **P** rule AND implies every PL(n) rule, for every value of $n \geq 2$.

Rule PL(n) says that if a collection of sentences is small enough (≤ n) and B nonmonotonically implies each of them, but B also implies-with-certainty that they cannot all hold, then B behaves like a "contradiction" in the sense that it implies every sentence. However, it is perfectly compatible with this rule that a "non-contradictory" sentence B may imply each of a large collection of jointly incompatible sentences, provided that collection consists of more than n distinct sentences.

In the case where n = 2, for instance, PL(2) requires that if $B|\sim A$ and $B|\sim\neg A$, then (since $B\cdot(A\cdot\neg A)|\sim \neg(A\cdot\neg A)$) it follows that $B|\sim D$ for every sentence D. More generally, rule PL(2) says that whenever $B\cdot(A_1\cdot A_2)|\sim \neg(A_1\cdot A_2)$, we cannot have both $B|\sim A_1$ and $B|\sim A_2$ unless B behaves like a contradiction (i.e. unless $B|\sim D$ for all D). Furthermore, each rule PL(n) for $n > 2$ implies this PL(2) rule.

Think of PL(n) this way. Consider the situation of the preface paradox (first raised by Makinson in [6]). The author's careful editing of his book strongly supports his belief that page i is error free, for each page i, but his knowledge of his own fallibility strongly implies that at least one error has slipped by in the editing process. Let each of the first $n-1$ sentences A_i be a sentence F_i that says that page i of the book is Free from error, and let sentence A_n be the sentence '$\neg(F_1\cdot\ldots\cdot F_{n-1})$', which says that not all $n-1$ pages of the book are error free – that at least one page contains an error. Notice that in this case the sentence $(A_1\cdot\ldots\cdot A_n)$ is the sentence '$(F_1\cdot\ldots\cdot F_{n-1}\cdot\neg(F_1\cdot\ldots\cdot F_{n-1}))$', which is an outright logical contradiction. So, given the author's knowledge B about the book, $B\cdot(F_1\cdot\ldots\cdot F_{n-1}\cdot\neg(F_1\cdot\ldots\cdot F_{n-1}))|\sim \neg(F_1\cdot\ldots\cdot F_{n-1}\cdot\neg(F_1\cdot\ldots\cdot F_{n-1}))$ simply follows from REFLEX and RW. Thus, the rule says that *when* the number of pages is $n-1$ (or fewer), B cannot consistently imply each of the $n-1$ claims that page i is error free and at the same time imply the claim that at least one of the pages contains an error. That is, rule PL(n) says that when the number of pages *is too small* ($n-1$ or smaller), B may imply each of these claims separately, and also imply that at least one of them is false, *only if* B itself is effectively a contradiction (in that B implies every claim, even it's own negation).

The "preface interpretation" of the A_i described here is merely an illustration of the rule. The same rule holds regardless of what the sentences A_i say. Notice too that this same rule, PL(n), also applies to a preface case for an n page book (and not merely to an n−1 page book, as in the above example) provided that that B *implies-with-certainty* that at least one page has an error – i.e. provided that $B\cdot(F_1\cdot...\cdot F_n) |\sim \neg(F_1\cdot...\cdot F_n)$.

It turns out that PL(n) is probabilistically sound for all and only the threshold levels $r > (n-1)/n$, as the next theorem shows.

Theorem 5: Probabilistic Soundness of Rule PL(n) for all and only the Threshold Values r > (n−1)/n. For $n \geq 2$, for each $r > (n-1)/n$ and each **CP** function P, the level-r consequence relation |~ corresponding to P (defined as 'B|~A' holds just in case $P[A \mid B] \geq r$) satisfies PL(n). Furthermore, for each $r \leq (n-1)/n$, there is a **CP** function P such that the level-r consequence relation |~ corresponding to P violates PL(n).

proof: To see that whenever $r > (n-1)/n$, PL(n) is satisfied by every probability function for threshold r: Suppose $r > (n-1)/n$ and $P[\neg(A_1\cdot\ldots\cdot A_n) \mid B\cdot(A_1\cdot\ldots\cdot A_n)] \geq r$, but there is a D such that $P[D \mid B] < 1$. (We show that for at least one of the A_i, $P[A_i \mid B] < r$.)

From the suppositions it follows that $0 = P[\neg(A_1\cdot\ldots\cdot A_n)\cdot(A_1\cdot\ldots\cdot A_n) \mid B] = P[\neg(A_1\cdot\ldots\cdot A_n) \mid B\cdot(A_1\cdot\ldots\cdot A_n)] \cdot P[(A_1\cdot\ldots\cdot A_n) \mid B] \geq r \cdot P[(A_1\cdot\ldots\cdot A_n) \mid B]$. So $P[(A_1\cdot\ldots\cdot A_n) \mid B] = 0$. Then $1 = P[\neg(A_1\cdot\ldots\cdot A_n) \mid B] = P[\neg A_1\vee...\vee\neg A_n \mid B] \leq P[\neg A_1 \mid B] +...+ P[\neg A_n \mid B] = (1 - P[A_1 \mid B]) +...+ (1 - P[A_n \mid B]) = n - (P[A_1 \mid B] +...+ P[A_n \mid B])$. So $P[A_1 \mid B] +...+ P[A_n \mid B] \leq (n-1)$. Now, given this, if $P[A_i \mid B] \geq r > (n-1)/n$ for every A_i, then we would have $(n-1) = n\cdot((n-1)/n) < n\cdot r \leq P[A_1 \mid B] +...+ P[A_n \mid B] \leq (n-1)$, contradiction!!! Thus, for one of the A_i, $P[A_i \mid B] < r$.

Conversely, to see that whenever $r \leq (n-1)/n$, rule PL(n) is violated by at least one consequence relation that corresponds to a conditional probability with threshold r, notice that there is clearly a probability function P with the following characteristics: for a sentence B such that $P[\neg B \mid B] < 1$ there are n sentences A_i such that $B \models (\neg A_1\vee...\vee\neg A_n)$, $B \models \neg(\neg A_i\cdot\neg A_j)$, and each $\neg A_i$ has the same probability given B. Then $P[\neg A_i \mid B] = 1/n$ for each i. So $P[A_i \mid B] = (n-1)/n \geq r$ for each i, yet there is a D such that $P[D \mid B] < 1$; and (since we also have that $B\cdot(A_1\cdot...\cdot A_n) \models (\neg A_1\vee...\vee\neg A_n)$) we have $r < 1 = P[(\neg A_1\vee...\vee\neg A_n) \mid B\cdot(A_1\cdot...\cdot A_n)] = P[\neg(A_1\cdot...\cdot A_n) \mid B\cdot(A_1\cdot...\cdot A_n)]$.

One additional observation is in order. Rule 6**P** is in effect the least upper bound of the PL(n) rules as n goes to infinity. This makes good sense in terms of the probabilistic models of these rules. A PL(n) rule corresponds to lower bound $(n-1)/n$ on the threshold in conditional-probabilistic models of consequence relations. As n increases, r is driven ever closer to 1, which is precisely the probabilistic threshold appropriate to AND.

For each $n \geq 2$, the LL(n+1) rule applies to any n+1 distinct sentences A_i that are implied-with-certainty by B to be mutually exclusive. Notice that LL(n+1) implies each LL(m) rules for $m \geq n+1$. So as n decreases the LL(n+1) rules become stronger.

Think of LL(n+1) this way. Consider a lottery (described by B) in which no two tickets can win. Let each of the n+1 sentences A_i say that ticket i will win, and suppose that B implies-with-certainty that no two ticket can win – i.e. that this lottery can have at most one winner. The expressions of form '$B\cdot(A_i\cdot A_j)|\!\sim \neg(A_i\cdot A_j)$' express this. Then, according to LL(n+1), for any given block of n+1 such tickets, B must nonmonotonically imply, for at least one ticket i, the claim that ticket i will not win. (Indeed, if B treats all n+1 tickets in the same way, then it must imply that each will not win – though the requirement that all tickets are treated equally is not a part of rule LL(n+1) itself.) The idea behind LL(n+1) is that if the number of tickets is *too large* (n+1 or bigger), and if B makes it certain that at most one can win, then at least one of the tickets must have such little chance of winning that B defeasibly implies that it won't win.[16]

There is no assumption here that the lottery is fair – that all tickets have the same chance of winning. So the logic only forces the issue for one of the n+1 tickets. Also notice that if there are more than (n+1) tickets, then for each block of (n+1) tickets, the rule applies. In other words, only for n or fewer tickets may B allow that each of them "might win" – i.e. only for n or fewer tickets may the conditional '$B|\!\sim \neg A_i$' fail to hold for each of them.

The "lottery interpretation" of the A_i here is, of course, merely an illustration of the rule. The same rule holds for all consequence relations in **Q**(n), regardless of how the A_i are interpreted.

It turns out that rule LL(n+1) is probabilistically sound for all and only the threshold levels $r \le n/(n+1)$, as the next theorem shows.

Theorem 6: Probabilistic Soundness of Rule LL(n+1) **for all and only the Threshold Values r ≤ n/(n+1).** For $n \ge 2$, for each $r > 0$ such that $r \le n/(n+1)$, and for each **CP** function P, the level-r consequence relation $|\!\sim$ corresponding to P (i.e. defined as '$B|\!\sim A$' holds just in case $P[A \mid B] \ge r$) satisfies LL(n+1). Furthermore, for each $r > n/(n+1)$, there is a **CP** function P such that the level-r consequence relation $|\!\sim$ corresponding to P violates LL(n+1).

proof: To see that whenever $0 < r \le n/(n+1)$, rule LL(n+1) is satisfied by every probability function applied to threshold r: Suppose $0 < r \le n/(n+1)$ and for each pair of the n+1 sentences A_i, $P[\neg(A_i\cdot A_j) \mid B\cdot(A_i\cdot A_j)] \ge r$. Notice that if for all D, $P[D \mid B] = 1$, then $P[\neg A_i \mid B] = 1 \ge r$ for each A_i, and we're done! So let's also suppose that for some D, $P[D \mid B] < 1$. (We want to show that for at least one A_i, $P[\neg A_i \mid B] \ge r$.)
From the suppositions it follows that for each distinct pair A_i and A_j, $0 = P[\neg(A_i\cdot A_j)\cdot(A_i\cdot A_j) \mid B] = P[\neg(A_i\cdot A_j) \mid B\cdot(A_i\cdot A_j)] \cdot P[(A_i\cdot A_j) \mid B] \ge r \cdot P[(A_i\cdot A_j) \mid B]$. So $P[(A_i\cdot A_j) \mid B] = 0$. Then we have $1 \ge P[A_1\vee...\vee A_{n+1} \mid B] = P[A_1 \mid B] +...+ P[A_{n+1} \mid B]$. Now, given this, if $P[A_i \mid B] > 1/(n+1)$ for every A_i, then we would have $1 \ge P[A_1 \mid B] +...+ P[A_{n+1} \mid B] > (n+1)\cdot(1/n+1) = 1$, contradiction!!! Thus, for at least one of the A_i, $P[A_i \mid B] \le 1/(n+1)$, so $P[\neg A_i \mid B] \ge n/(n+1) \ge r$.
Conversely, to see that whenever $r > n/(n+1)$ rule LL(n+1) is violated by at least one consequence relation that corresponds to a conditional probability function with threshold r, notice that there is clearly a probability function P with the following characteristics: for a sentence B such that $P[\neg B \mid B] < 1$, there are n+1 sentences A_i such that $B \models (A_1\vee...\vee A_{n+1})$, $B \models \neg(A_1\cdot A_j)$ for each pair, each A_i has

[16] Kyburg first raised the lottery paradox in [9], and treated it further in [10].

the same probability given B. Then for $r > n/(n+1)$ we have $P[A_i \mid B] = 1/(n+1)$ for each A_i, so $P[\neg A_i \mid B] = n/(n+1) < r$ for each A_i. But $B \cdot (A_1 \cdot A_j) \models \neg(A_1 \cdot A_j)$, so $P[\neg(A_i \cdot A_j) \mid B \cdot (A_i \cdot A_j)] = 1 \geq r$ for each pair.

Rule LL(n+1) is clearly compatible with rule PL(n), since the probability above r part of every conditional probability function satisfies both rules whenever $(n-1)/n < r \leq n/(n+1)$. Thus, Theorems 5 and 6 show that the rules for **Q**(n) are probabilistically sound for precisely those thresholds r greater than $(n-1)/n$ but no greater than $n/(n+1)$.

The logic of the *Rational Consequence Relations*, **R**, may reasonably be called **Q**(∞). For one thing, as n grows ever larger, the bounds r such that $(n-1)/n < r \leq n/(n+1)$ for probabilistic models of the **Q**(n) logics approach 1. For another thing, the rules PL(n) are implied by AND, and are superseded by it when r = 1. Furthermore, as n increases, rules LL(n+1) approaches vacuity. At the same time, at r = 1 rule NR (Negation Rationality) becomes too weak, and rule RM (Rational Monotony) supersedes it. Finally, at r = 1 the rules of **R** are probabilistically sound and complete – i.e., the probability 1 part of each conditional probability function constitutes a consequence relation in **R**, and each consequence relation in **R** is the probability 1 part of some conditional probability function.

6 Concluding Remarks

We've seen that the rules for **Q**(n) are probabilistically sound for precisely those thresholds r above $(n-1)/n$ and no greater than $n/(n+1)$. For any threshold level in this interval, the part of each conditional probability function above that level constitutes a consequence relation in **Q**(n). And for each threshold level outside of this interval, there is a conditional probability function whose part above that threshold constitutes a consequence relation not in **Q**(n). However, the rules of **Q**(n) have not been shown to be probabilistically complete. Indeed, in a private communication David Makinson has shown me that there are (linear ranked) consequence relations satisfying **Q**(n)'s rules that are not probabilistically modelable at any threshold level. So although in the extreme case **Q**(∞), i.e. **R**, we do have probabilistic completeness, additional rules are needed to restrict the class of **Q**(n) consequence relations for $2 \leq n < \infty$ to the probabilistically modelable ones. What are these additional rules like?

Those of you who are familiar with the logic of qualitative probability (a.k.a. comparative probability) know that one way to get probabilistically modelable qualitative probabilities is to introduce a rule that says, in effect, that each qualitative probability relation is extendable to a relation on language that includes sentences that form arbitrarily fine partitions, where all sentences of a given partition are "approximately qualitatively equal". It may then be shown that these equal partition sentences can be used as a standard of comparison to fix numerical probabilistic weights on all sentences of the language, and thereby generate a numerical probability function. A rather similar idea may be applicable to the consequence relations in **Q**(n). I investigated one way to make this idea work in an earlier article (see [2], sections 4.2 and 4.3). But the presuppositions of that approach seem overly strong. So, the issue of how to plausibly supplement **Q**(n) in a way that completely characterizes probabilistic consequence relations remains an open question.[17]

Appendix

Theorem. Given any threshold r such that $0 < r < 1$, for each rule OR, CM, and AND, there is a **CP** function P (indeed, there is a classical Kolmogorov probability function) that violates the rule. However, for threshold $r = 1$ each of these rules is probabilistically sound with respect to the **CP** functions.

Proof: Let r be any fixed real number such that $0 < r < 1$. To show that a given rule fails for r we only need show that there is a probability function P such that when we replace each expression of form 'Z |~ Y' in the rule by '$P[Y \mid Z] \geq r$' the rule fails for P. (Notice that in each case below the probabilistic model P that violates the rule can be one of the usual Kolmogorov probability functions; it need not be one of the "non-standard" functions in **CP**.) Furthermore, we show that replacing each expression of form 'Z |~ Y' in the rule by '$P[Y \mid Z] = 1$' yields a theorem of **CP**-probability.

1. OR: We want to show that for each r such that $0 < r < 1$, there is a function P and sentences C, B, and A such that $P[A \mid B] \geq r$ and $P[A \mid C] \geq r$ but $P[A \mid B \vee C] < r$.

Clearly for any r such that $0 < r < 1$, there is a P such that for some C, B, and A, the following hold: $P[B \cdot C \mid C \vee B] = P[\neg B \cdot C \mid C \vee B] = P[B \cdot \neg C \mid C \vee B] = 1/3$, and $P[A \mid B \cdot C] = r+\varepsilon$, while $P[A \mid \neg B \cdot C] = P[A \mid B \cdot \neg C] = r-\varepsilon$, for some $\varepsilon > 0$ but small enough that $0 < r-\varepsilon < r+\varepsilon < 1$.

Now, $P[B \mid C] = P[B \mid C \cdot (B \vee C)] = P[B \cdot C \mid B \vee C] / P[C \mid B \vee C] = P[B \cdot C \mid B \vee C] / (P[B \cdot C \mid B \vee C] + P[\neg B \cdot C \mid B \vee C]) = 1/2$; and so also $P[\neg B \mid C] = 1/2$. So, $P[A \mid C] = P[A \mid B \cdot C] \cdot P[B \mid C] + P[A \mid \neg B \cdot C] \cdot P[\neg B \mid C] = (1/2) \cdot (r+\varepsilon + r-\varepsilon) = r$.

Similarly, $P[C \mid B] = 1/2$ and $P[\neg C \mid B] = 1/2$; so $P[A \mid B] = r$.

But since $P[\neg B \cdot \neg C \mid C \vee B] = 0$, $P[A \mid C \vee B] = P[A \mid B \cdot C] \cdot P[B \cdot C \mid C \vee B] + P[A \mid B \cdot \neg C] \cdot P[B \cdot \neg C \mid C \vee B] + P[A \mid \neg B \cdot C] \cdot P[B \cdot \neg C \mid C \vee B] = (1/3) \cdot (r+\varepsilon + r-\varepsilon + r-\varepsilon) = r - \varepsilon/3 < r$.

For $r = 1$: suppose $P[A \mid B] = P[A \mid C] = 1$. Then (unless for all D, $P[D \mid B \vee C] = 1$, and we're done) $P[A \mid B \vee C] = P[(A \cdot B) \vee (A \cdot C) \mid B \vee C] = P[A \cdot B \mid B \vee C] + P[A \cdot C \mid B \vee C] = P[A \mid B] \cdot P[B \mid B \vee C] + P[A \mid C] \cdot P[C \mid B \vee C] = P[B \mid B \vee C] + P[C \mid B \vee C] = P[B \vee C \mid B \vee C] = 1$.

2. CM: We want to show that for each r such that $0 < r < 1$, there is a function P and sentences C, B, and A such that $P[B \mid C] \geq r$ and $P[A \mid C] \geq r$ but $P[A \mid C \cdot B] < r$. We divide the proof into two cases, depending on whether $r \leq 1/2$ or $r > 1/2$.

Case 1: First, suppose that $0 < r \leq 1/2$. Clearly there is a P such that for some C, B, and A, the following hold: $P[A \cdot B \mid C] = P[\neg A \cdot \neg B \mid C] = 0$, $P(\neg A \cdot B \mid C] = r$, $P(A \cdot \neg B \mid$

[17] I am indebted to Greg Wheeler, two conference referees, and the participants at CMSRA-IV (Lisbon, September 22-23, 2005) for many valuable comments and suggestions. Special thanks to David Makinson for a number of stimulating email communications about these logics.

C] = 1−r ≥ r. Then P[B | C] = P[¬A·B | C] = r; P[A | C] = P[A·¬B | C] = 1−r ≥ r. However, P[A | C·B] = P[A·B | C] / P[B | C] = 0 < r.

Case 2: Alternatively, suppose that 1 > r > 1/2. There is a P such that for some C, B, and A, the following hold: P[A·B | C] · $[(1-r)/r^2]$ = P[¬A·B | C] = P[A·¬B | C] > 0 and P[¬A·¬B | C] = 0. We need to check that the function P as just specified is not "over-constrained". First, 1 = P[A∨¬A | C] = P[A·B | C] + P[A·¬B | C] + P[¬A·B | C] + P[¬A·¬B | C] = P[A·B | C] · $(1 + 2\cdot[(1-r)/r^2])$ = P[A·B | C] · $[(r^2 - 2r + 1) + 1]/r^2$ = P[A·B | C] · $[(1-r)^2 + 1]/r^2$; so our specification of P only requires that P[A·B | C] = $r^2/[(1-r)^2 + 1]$, which for 1/2 < r ≤ 1 is clearly greater than 0, and cannot be greater than 1 (else: $r^2 > (1-r)^2 + 1 = 2 - 2r + r^2$, so 2r > 2, which is impossible since r ≤ 1). And since P[A·B | C] = $r^2/[(1-r)^2 + 1]$, we have P[¬A·B | C] = P[A·¬B | C] = P[A·B | C] · $[(1-r)/r^2] = (1-r)/[(1-r)^2 + 1]$, which for 1 > r > 1/2 is clearly between 0 and 1.

Now we only need check that P[A | C] ≥ r and P[B | C] ≥ r, but P[A | C·B] < r. P[A | C] = P[A·B | C] + P[A·¬B | C] = $r^2/[(1-r)^2 + 1] + (1-r)/[(1-r)^2 + 1] = [(1-r)^2 + r]/[(1-r)^2 + 1]$, which must be ≥ r (else: $r > [(1-r)^2 + r]/[(1-r)^2 + 1]$, so $r\cdot(1-r)^2 + r > (1-r)^2 + r$, so r > 1).

Similarly, P[B | C] = P[A·B | C] + P[¬A·B | C] = $[(1-r)^2 + r]/[(1-r)^2 + 1] \geq r$.

But, P[A | C·B] = P[A·B | C] / P[B | C] = $(r^2/[(1-r)^2 + 1])/([(1-r)^2 + r]/[(1-r)^2 + 1])$ = $r^2/[(1-r)^2 + r]$, which is must be < r (else: $r \leq r^2/[(1-r)^2 + r]$, so $[(1-r)^2 + r] \leq r$, so $(1-r)^2 \leq 0$, which cannot be for r < 1.)

For r = 1: suppose P[A | C] = P[B | C] = 1. Then *either* for all D, P[D | C] = 1, so 1 = P[A·B | C] = P[A | B·C]·P[B | C] = P[A | C·B] *or else* 1 = P[A | C] = P[A·B | C] + P[A·¬B | C] = P[A | B·C]·P[B | C] + P[A | ¬B·C]·P[¬B | C] = P[A | C·B].

3. AND: We want to show that for each r such that 0 < r < 1, there is a function P and sentences C, B, and A such that P[B | C] ≥ r and P[A | C] ≥ r but P[B·A | C] < r. Choose any such r. Clearly there exists a P such that for some C, B, and A, P[B | C] = r and P[A | C] = r, and where B and A are independent for P given C, so that P[B·A | C] = P[B | C] · P[A | C] = $r^2 < r$.

For r = 1: suppose P[B | C] = 1 and P[A | C] = 1. Then *either* for all D, P[D | C] = 1, so P[A·B | C] = 1 *or else* 1 = P[A | C] = P[A·B | C] + P[A·¬B | C] = P[A·B | C] + P[A | ¬B·C]·P[¬B | C] = P[A·B | C].

References

1. Arlo-Costa, H., Parikh, R.: Conditional Probability and Defeasible Inference. Journal of Philosophical Logic **34** (2005) 97-119
2. Hawthorne, J.: On the Logic of Nonmonotonic Conditionals and Conditional Probabilities. Journal of Philosophical Logic **25** (1996) 185-218
3. Hawthorne, J.: On the Logic of Nonmonotonic Conditionals and Conditional Probabilities: Predicate Logic. Journal of Philosophical Logic **27** (1998) 1-34
4. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic Reasoning, Preferential Models and Cumulative Logics. Artificial Intelligence **44** (1990) 167-207
5. Lehmann, D., Magidor, M.: What Does a Conditional Knowledge Base Entail? Artificial Intelligence **55** (1992) 1-60
6. Makinson, D.: The Paradox of the Preface. Analysis **25** (1965) 205-207

7. Makinson, D.: General Patterns in Nonmonotonic Reasoning. In Gabbay, D., Hogger, C., and Robinson, J. (eds.): Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3. Oxford University Press (1994). 35-110
8. McGee, V.: Learning the Impossible. In Eells, E. and Skyrms, B. (eds.): Probability and Conditionals: Belief Revision and Rational Decision. Cambridge University Press, Cambridge (1994).
9. Kyburg, H.: Probability and the Logic of Rational Belief. Wesleyan University Press, Middletown (1961).
10. Kyburg, H.: Conjunctivitis. In: Swain, M. (ed.): Induction, Acceptance, and Rational Belief. Reidel, Dordrecht (1970).