A Formal Analysis of Relevance

James P. Delgrande
School of Computing Science,
Simon Fraser University,
Burnaby, B.C.,
Canada V5A 1S6

Jeff Pelletier
Department of Philosophy,
University of Alberta,
Edmonton, Alberta
Canada T6G 2H1

Abstract

We investigate the notion of *relevance* as it pertains to 'commonsense', subjunctive conditionals. Relevance is taken here as a relation between a property (such as having a broken wing) and a conditional (such as birds typically fly). Specifically, we explore a notion of 'causative' relevance, distinct from 'evidential' relevance found, for example, in probabilistic approaches. A series of postulates characterising a minimal, parsimonious concept of relevance is developed. Along the way we argue that no purely logical account of relevance (even at the metalevel) is possible. Finally, and with minimal restrictions, an explicit definition that agrees with the postulates is given.

1 Introduction

The notion of a *relevant* property would seem to be intuitively clear: naïvely, a property may be taken as relevant to some other property if it leads us to change our mind concerning whether the second property holds. However we claim that this is not the case, and that when looked at a bit more closely, various issues and distinctions arise almost immediately. Consider the following assertions.

1a. Birds (normally) fly.

b. Green birds fly.

c. Birds with broken wings do not fly.

d. Green birds with broken wings do not fly.

Intuitively, being green is irrelevant to whether or not a bird flies, whereas having a broken wing is relevant. However, while having a broken wing is relevant to whether a green bird flies, it is not apparent that having a broken wing *and* being green is relevant to a bird's being able to fly – at least, if it is relevant, it is not particularly parsimonious.

Such a notion of relevance is found in a wide variety of settings, including counterfactual statements, deontic assertions, hypothetical assertions, and other subjunctive conditionals. Consider the following counterfactual assertions, wherein a lively party took place the previous night [Lew73]:

2a. If John had gone to the party, it would have been a dull party.

b. If John and Susan had gone to the party, it would have been a lively party.

As [2a] shows, John's individual presence is relevant to the outcome of the party; but in the case of John and Susan's joint presence, one might argue either way. On the one hand, neither John nor Susan had in fact gone to the party, and it was a lively affair. If both had gone, the outcome would be the same and so their joint presence is (arguably) irrelevant. Alternately it might be pointed out that even though the outcome is the same, in the second case a lively party results for different reasons than the first, and so their presence is relevant.

Consider further the deontic assertions:

3a. One should drive within the speed limit.

b. One may, in certain emergencies, exceed the speed limit.

Clearly there *are* grounds, or relevant conditions, for exceeding the speed limit. Similarly, in the following hypothetical assertions, relevant conditions play a role in a planning task.

4a. If the dentist's office is above the third floor, I will take the elevator.

b. If the dentist's office is above the third floor but the elevator is broken,
then I will take the stairs.

Our intent in this paper is to formally investigate the concept of relevance as given in the preceding examples. Clearly, this conception of relevance is more general than that of *conditional independence* or of related notions, which are found in approaches based on probability theory or in approaches which (one way or another) involve certainty. Of the examples discussed above, only [1] admits an obvious probabilistic interpretation. The examples concerning when it may be permissible to speed ([3]), and how I might get to my dentist's office ([4]), clearly involve a notion of relevance but in a setting where probability theory plays no role.

However, it is not immediately clear what the scope of our investigation will be, nor indeed what *form* our account of relevance should take. Consequently, in Section 2, we begin by reviewing previous approaches to defining and dealing with relevance. In Section 3 we first examine the form that a definition of relevance should take. We argue that a definition at the metalevel is required. In addition, we feel that 'relevant' is a concept for which we have no deep understanding in the context under discussion; consequently we suggest that an implicit definition of the term is most appropriate. Next, we briefly introduce the formal framework with respect to which our definition is framed, that of conditional logics. Following this we discuss the form of the relation 'relevant'. First, we consider where relevant is a binary relation on propositions; for example one might say that being fit is relevant to being in good health. Against this we argue that relevance is most appropriately expressed as a relation between a proposition (such as having a broken wing) and a proposition that corresponds to a 'weak' conditional (such as birds typically fly).

Informally, our interests lie with what might be called 'causative relevance'. Thus we say that being an albino is relevant to a raven's being black, since ravens are normally black, whereas albino ravens are normally not black. Being non-albino is not relevant, since normal non-albino ravens are also black. This can be contrasted with probabilistic approaches, where both being albino and being non-albino are considered relevant, because each piece of information would help in predicting a raven's colour. However, as we argue subsequently, the differences run significantly deeper.

In Section 4 we explore this notion of relevance by proposing a series of postulates (and non-postulates) implicitly characterising this term. In so doing, we distinguish several senses of the term. We discuss there what might be called 'promiscuous' relevance, which conforms to the principle:

AUG: If α is relevant to β being a γ then $\alpha \wedge \delta$ is relevant to β being a γ.

We argue that promiscuous relevance is not a reasonable relation for the cases we are considering, since it makes too many things relevant. Rather, we are interested in a notion of *minimal* or *parsimonious* relevance. Moreover, in the conclusion we suggest that a notion of relevance is crucial in defeasible reasoning, but that any definition employing **AUG** will be too broad to be useful.

We also discuss a more subtle distinction within types of relevance, between what we call *that-relevance* and *whether-relevance*. The difference is between whether a property is relevant for something having a property (say, a raven being black), versus whether-or-not the individual has the property (such as is found in probabilistic approaches).

In any case, the set of postulates developed in Section 4 are suggested as comprising a set of conditions that notions of (non-promiscuous) relevance ought conform to. Hence these postulates comprise a base, implicit definition. However we also argue that no logical, language-independent account of relevance can be wholly adequate; rather some domain-specific or syntactic constraints must also be adopted. We subsequently suggest a restriction, whereby the set of potentially relevant conditions is denoted by the set of literals, or their conjunctions, in a given theory. Using this we propose a specific explicit definition for the term *relevant*, and show that this explicit definition conforms to our set of postulates. Section 5 contains a discussion, while Section 6 gives concluding remarks and suggestions for further research. Proofs of theorems are found in the appendix.

2 Related Work

An informal notion of relevance occurs in mechanical theorem proving wherein one would want to use information relevant to the task of establishing or refuting some goal. In resolution theorem proving (see [Lov78] or [Nil80]) for example, in the *set-of-support strategy*, clauses for resolving upon are selected in part from those resulting from the negation of the goal and their descendants. Since clauses are selected that bear directly on a goal, only clauses relevant to the goal are resolved upon. However the focus in the theorem proving arena is on domain-independent strategies, and these strategies rely solely on the syntactic form of the clauses. Moreover, in fact no account of relevance is given. Rather strategies are justified (at the metalevel) by *appeal* to a pretheoretic notion of relevance. That is, the view taken amounts to having some notion in advance as to what will 'help' a proof, and then labelling that as 'relevance.' This general attitude is also held by [LS93], and we reject it in their application also (see below). Clearly then such approaches are not what we are interested in and will not help in the present endeavour.

For approaches that explicitly address relevance, we can distinguish those where a relevant property alters the *probability* (degree of acceptance, etc.) attached to a statement, from those where a relevant property alters the *acceptance* of a statement. It should be clear that these are distinct categories and that the latter is not simply a restricted version of the former. The example of the dentist's office ([4] above) for example, clearly does not involve probabilities. Even approaches which can be seen as involving a probabilistic interpretation ('birds typically fly' for example) can also be given a non-probabilistic interpretation [Del87].

2.1 Quantitative Approaches

In classical probability theory, propositions p and r are *conditionally independent*, given the information or *evidence* e if

$$P(r \mid p \wedge e) = P(r \mid e) \text{ whenever } P(p \wedge e) \neq 0.$$

Naïvely, we might equate this concept of conditional independence with that of *irrelevance*, and define a proposition to be relevant if and only if it is not independent. In [Pea88, p. 84] a set of axioms characterising conditional independence is given. [DP94] (see below, Section 2.2) gives a definition of independence in databases compatible with these axioms.

Peter Gärdenfors [Gär78], building on earlier work by Keynes and Carnap, has argued that this definition, taken as a definition of (ir)relevance, is overly weak. He gives a series of postulates governing relevance and, following these postulates, the following explicit definition is proposed:

Definition 2.1

1. *p is* irrelevant *to r on evidence e iff either*
   (a) $P(r \mid p \wedge e) = P(r \mid e)$ *and for all sentences q, if* $P(r \mid q \wedge e) = P(r \mid e)$ *and* $P(p \wedge q \wedge e) \neq 0$ *then* $P(r \mid p \wedge q \wedge e) = P(r \mid e)$*, or*
   (b) $P(p \mid e) = 0$.
2. *p is* relevant *to r on e iff p is not irrelevant to r on e.*

Clause (1a) states that if p is irrelevant to r, then if p were to be included in the evidence e, then no other sentences irrelevant to r and e should become relevant to r on this new evidence. This approach is discussed further in Section 5 below. See also [Sch86], where various modifications of this definition are proposed and the connection between relevance and (scientific) confirmation is emphasised.

Two approaches to using (as opposed to *defining*) relevance in a probabilistic setting are discussed and compared in [Pea88]. First, *Bayesian networks* (which are directed acyclic graphs where nodes represent variables and edges designate probabilistic dependencies) can be regarded as a means of implementing (and determining) new dependence relationships, that is, relevancies, in a network. Second, the *maximum entropy* approach is a mechanism for selecting probability distributions that minimise dependencies relative to a background theory; as such it can be seen as implementing a notion of relevance. Both approaches involve restrictions: In Bayesian networks, nodes represent random variables; hence there is no notion of compound formulas. Maximum entropy fails to deal with causal relations; in addition, inferences are dependent on how the database is expressed.

[DFHP94] examine various notions of (in)dependence with respect to possibility theory. In contrast to related work, dependence here is given as a binary relation, and so one would have 'B and C are independent' rather than the more common 'B and C are independent given A'. In Section 3.3.1 we argue that such an approach is unsuitable for a full account of relevance.

2.2 Qualitative Approaches

The seminal work in this area is [Goo61], which addresses what it means for a predicate to be *about* some other predicate. Two notions of 'aboutness', a binary and a ternary relation, are given. These correspond to concepts of 'absolutely about' and 'relatively about'. In the first case we might say that the statement 'Ravens are black' is absolutely about black things, as well as ravens. In the second case, we can say that 'York County grows corn' is relatively about Ontario, relative to the statement 'York County is in Ontario'. This latter notion is discussed further in Section 5 below. In [Goo73, I.2] the problem of relevant conditions with respect to counterfactual conditionals is discussed as part of a wider discussion of counterfactuals.

The area of relevance logic [AB75] can be seen as dealing with a binary notion of relevance. Relevance logic grew out of a dissatisfaction with material implication as expressing a plausible notion of 'if ... then ...'. Roughly, the feeling was that for a conditional $\alpha \rightarrow \beta$ to hold, the antecedent should be relevant to the consequent. In these logics the connective $\rightarrow$ is strictly weaker than the material conditional, specifically as a consequence of this requirement. Hence we can regard each of these logics as formalising some notion of 'relevant', and we can assert, for example, that α is relevant to β relative to theory $\mathcal{T}$ in logic R iff $\mathcal{T} \vdash_R \alpha \rightarrow \beta$. This notion then (arguably) has the restricted interpretation 'deductively-relevant-to'.

In Artificial Intelligence, the problem of irrelevance in problem-solving systems is addressed in [SG87]. There, the goal is to explicate the notion of a fact f being irrelevant to fact g given a knowledge base M; such information would be used to prune the search space of some problem. M is some arbitrary set of *sentences*. Essentially, f is *irrelevant* to g with respect to theory M if there exists a weakening of M, M', such that M' is non-committal on f, and M' is maximal. That is,

1. $M' \subseteq M$,
2. $M' \models g$ (and so $M \models g$), but
3. $M' \not\models f$ and $M' \not\models \neg f$, and
4. M' is the maximal set satisfying the above properties.

This approach provides a means of reasoning about irrelevance: beginning with M we determine, in the metatheory, which sentences are irrelevant to which others. The authors provide a partial axiomatisation, or list of properties that characterise this notion, and so an implicit definition of relevance.

This definition is syntactic, being stated in terms of the proofs of a statement in a database, and depends crucially on the *form* of the sentences in the knowledge base. If we were instead to model knowledge bases semantically, by their deductive closure for example, we would have $M = Cn(M)$ and $M' = Cn(M')$. In this case the definition yields that f is irrelevant to g iff $g \not\models f$ and $g \not\models \neg f$. That is, regardless of the knowledge base M, one sentence is relevant to another exactly when either the latter implies the former or it implies the former's negation. And if we were to add the condition that one sentence is irrelevant to a second only if the second is irrelevant to the first, we get the result that two sentences are relevant to one another exactly in case they are logically equivalent. Thus the definition becomes trivial for a semantic characterisation of a knowledge base.[1]

Levy and Sagiv [LS93] follow up and generalise this direction of research. They distinguish weak from strong irrelevance of a fact f to a query about g given a knowledge base M, on the basis of whether f is absent from some or all derivations of g from M, respectively. Their interests are quite limited, too much so for our present purposes. As they say (p.139), "our analysis is not an attempt to formalise the commonsense notion of irrelevance or argue for properties of such a notion .... Our goal is to utilise the notion of irrelevance ..."

[DP94] addresses causality in logical reasoning and, in so doing, presents a qualitative, logical notion of conditional independence. In the following definition X, Y, and Z are disjoint sets of atomic propositions. A conjunction of all literals from X is denoted $\hat{X}$.

Definition 2.2 *A database Δ finds X* independent of Y given Z, *just if the consistency of*

$$\Delta \cup \{\hat{Z}, \hat{X}\} \text{ and } \Delta \cup \{\hat{Z}, \hat{Y}\} \text{ implies the consistency of } \Delta \cup \{\hat{Z}, \hat{Y}, \hat{X}\}.$$

That is, if Δ has information about Z, then the addition of information about Y to Δ will not change the database's belief in any information about X. This definition only applies to (disjoint) sets of atomic propositions, and so has nothing to say about the independence of arbitrary sentences. In addition (see Section 5), this definition supports arbitrary augmentation, and so does not capture our requirement of parsimony.

Two other approaches to using relevance can also be noted here. First [Pea90] describes a conditional closure operation based on rankings of models of a set of conditionals, where the ranking depends only on which conditionals are verified or falsified. Irrelevant properties take no part in the ranking and so, in the end, play no role in the inference relation. Although straightforward, this approach has a number of shortcomings, including failure to allow inheritance of properties. This in turn means that one cannot appeal to the inference relation to supply an adequate, implicit definition of relevance.

[1]Unfortunately too, the axiomatisation provided is inconsistent. Axiom 7 states that for $g \in M$, and $f \not\equiv g$, f is irrelevant to g in M; so A is irrelevant to $A \wedge B$. Axiom 3 states that if A is irrelevant to $A \wedge B$ then A is irrelevant to (the individual conjuncts) A and B. But Axiom 4 states that no sentence is irrelevant to itself.

Finally, a common mechanism in Artificial Intelligence for dealing with exceptional individuals and classes is by means of circumscribing 'abnormality' predicates [McC86]. We can express in first-order logic that birds that are not abnormal fly as follows:

$$\forall x Bird(x) \wedge \neg Ab(x) \implies Fly(x).$$

Circumscribing *Ab* amounts to asserting that the number of things that *Ab* holds for is minimal. In doing this circumscribing we in essence move to the metalevel to solve a circumscription schema (a second-order formula) and then, if the result is first-order expressible, we reincorporate it into the object level. Hence circumscription may be viewed as a sort of metalevel 'theory augmentation' operation, since we compute the circumscription at the meta-level, but we then augment our object-level theory with the result of this computation.

3 The Formal Framework

This section examines the overall framework within which a definition for relevance will be phrased. In the first subsection we briefly discuss potential *forms* of definitions, and discuss which form might be suitable for 'relevant'. In the second subsection we introduce the specific formal framework with respect to which the definition will be formulated, that of conditional logics. The third subsection examines the question of what it is exactly that relevance is a relation among.

3.1 Forms of Definitions

There are various forms that a definition can take. For our purposes, there are two basic choices that need to be made. First there is the issue of whether the definition is a term in the *object language* or in the *metalanguage*. And second is the issue of whether the definition is to be *explicit* or *implicit*.

The first of these distinctions is between being a term *in* a theory versus being a term *describing* a theory. Hence 'α is relevant to γ' is a formula of the object language in the first conception but not in the metalinguistic conception. In the approach discussed here we take relevance to be a term in the metalanguage, as indeed do the approaches discussed in Section 2. That is, there is general agreement that relevance irreducibly involves or rests on logical consequence. Thus for example, Goodman [Goo61] explicitly states that '*aboutness* <is> defined in terms of certain logical consequences', and goes on to say that this notion will depend in part on what logic is presupposed. Similarly, in our account, we appeal to what is entailed and what is not entailed in a given theory. But this then means that our 'relevance' (and Goodman's 'aboutness') will not be stated as formulas *in* the language but rather will be statements describing what entailments the theory has.

This commitment to the metalevel has various consequences. For one thing, the notion of reiteration or self-embedding of 'relevant' is problematic, whereas there seems to be no problem in an object language version to say 'α is relevant to γ's being relevant to β', and the like. Furthermore, since the metatheoretic definition presupposes that 'relevant' will not be a term in the object language, it follows that no metatheoretic definition can be of direct use in the object language. So the existence of such a definition will not help in determining whether an argument is or isn't valid – the train of explanation instead goes the other way: knowledge of which arguments are valid or invalid is what allows us to define relevance.[2]

The second of the considerations is whether a definition is implicit or explicit. The difficulty with having only implicit definitions is that one doesn't in fact know what the term designates – only that whatever it designates will obey these rules. One has no assurances as to how many *distinct* things there are which obey the axioms.[3] Not only do we hold that there can be many different underlying logics with respect to which one might define relevance, but we also take the position that there may not be a single definition of the term even within the logic that is chosen. Rather, there may be a *family* of relevance relations within that logic, in much the same way as in knowledge base revisions [Gär88] there is a family of possible revision operations. Consequently, in the sequel we provide a set of postulates which should govern *any* account of relevance. We also indicate divergent intuitions concerning this notion and provide postulates for specialisations of this notion.

3.2 Conditional Logic

A common feature of the introductory examples is that they may be consistently expressed using conditionals in the form $\alpha \Rightarrow \gamma$ yet $\alpha \wedge \beta \Rightarrow \neg\gamma$. That is, as the antecedent of a conditional is strengthened, the consequent may change to its negation, and both these conditionals may be simultaneously and non-trivially satisfied. Clearly, this pattern could be continued arbitrarily: we might agree that birds with well-splinted broken wings can fly, or that if John, Susan, and Chris had gone to the party it would have been dull. Equally clearly, the connective in these examples cannot be classical material implication, strict (modal) implication, or logical entailment.

It is also not difficult to come up with examples that violate transitivity of the 'conditional'. So, penguins are birds, and birds (normally) fly, but penguins don't (normally) fly.[4] Similarly, we might accept the truth of the assertions 'John would be happier if he worked less' and 'if John broke his leg, he would work less' but not 'John would be happier if he broke his leg'.

The class of *conditional logics* [Sta68, Lew73, Nut80] has been proposed as a formal approach to capture such weak (subjunctive) conditionals. The semantic theory for these logics is usually expressed using a possible worlds semantics where worlds are arranged in some order (generally at least a preorder). The general idea is that a conditional $\alpha \Rightarrow \beta$ is true at a world just when, in the least subset S of worlds in which α is true, β is true in all of S. In this sense the conditional can be regarded as a necessity operator on β where the

[2]However, a metatheoretic definition could play an indirect role in reasoning at the object level, for example in nonmonotonic reasoning. This possibility is discussed in Section 6.

[3]The threat is that there can be an infinity of different, non-intended models of the axioms.

[4]We should point out that although our formulas are all stated as propositional formulas, we have in mind the first-order version where all these formulas are universally quantified. Here we merely leave the wide-scope quantifiers implicit, so that 'birds fly' is symbolised as $B \Rightarrow F$, but is intended as $\forall x(B(x) \Rightarrow F(x))$. The semantics we provide below is for the propositional case; an extension to the first-order case can be found in [Del87]. Note though there is much disagreement as to the best way to carry out this first-order extension in conditional logics.

subset of 'pertinent' worlds depends on α; for this reason the operator $\Rightarrow$ is referred to as a *variably strict conditional* or simply a *variable conditional*. Thus in Lewis's central approach, $\alpha \Rightarrow \beta$ is true if those worlds in the set ('sphere') of worlds most like our own that have α true, all have β true. (I.e., in this sphere no world has $\alpha \implies \beta$ false.) In these logics the sets of sentences

$$\{\alpha \Rightarrow \gamma, \ \ \alpha \wedge \beta \Rightarrow \neg\gamma\} \ \text{ and } \ \{\alpha \Rightarrow \beta, \ \ \beta \Rightarrow \gamma, \ \ \alpha \Rightarrow \neg\gamma\}$$

may be simultaneously and nontrivially satisfied. For this reason we also refer to the conditional $\Rightarrow$ as a *weak* conditional. There are of course major differences between conditional logics intended for differing applications. For example, counterfactual reasoning deals with conditionals where, in the interesting case, the antecedent is (contingently) false but where the conditional as a whole may be either true or false. For a logic of default properties, the interesting case is where the antecedent of the conditional is contingently true.

In the following, we will take the language of the logic to be that of (standard) propositional logic (PC) augmented with a binary operator $\Rightarrow$, reserving $\implies$ for material implication. Our development generally follows [Bur81]. Sentences are interpreted in terms of a *model* $\mathcal{M} = \langle W, R, P\rangle$ where:

1. W is a set (of worlds),
2. R is a ternary accessibility relation on worlds, and
3. P maps atomic sentences onto subsets of W.

$R_w w_1 w_2$ has the informal reading 'according to world w, w_2 is accessible from w_1.' For all intents and purposes we will treat R as a binary relation on its last two arguments; thus the first argument is subscripted, as a reminder that we are (essentially) dealing with a binary relation, but indexed by a particular world. For conditional logics, regardless of application, the relation (on the last two arguments) is a preorder; hence we assume that R is reflexive and transitive. For simplicity, we adopt the Limit Assumption [Lew73], that the order on worlds is well-founded. The absence of this assumption would not change the exposition following; it would however slightly complicate the formal details.

We define:

$$W_w = \{w_1 \mid \text{ for some } w_2 \in W, R_w w_1 w_2\}.$$

Further we require that R satisfy the following requirements:

Reflexivity:

$$\forall w \in W, \forall w_1 \in W_w, R_w w_1 w_1.$$

Transitivity:

$$\forall w \in W, \forall w_1, w_2, w_3 \in W_w, \text{ if } R_w w_1 w_2 \text{ and } R_w w_2 w_3 \text{ then } R_w w_1 w_3.$$

Truth is defined as expected, with the interesting case being for the variable conditional:

Definition 3.1

1.

$$\mathcal{M}, w \models p \ \ \textit{for} \ \ p \in \mathbf{P} \ \ \textit{iff} \ \ w \in P(p).$$

2.

$$\mathcal{M}, w \models \neg\alpha \ \ \textit{iff} \ \ \textit{not} \ \mathcal{M}, w \models \alpha.$$

3.

$$\mathcal{M}, w \models \alpha \wedge \beta \ \ \textit{iff} \ \ \mathcal{M}, w \models \alpha \ \ \textit{and} \ \ \mathcal{M}, w \models \beta.$$

4.

$$\mathcal{M}, w \models \alpha \Rightarrow \beta \ \ \textit{iff} \ \ \textit{for every } w_1 \in W_w \textit{ such that } \models^{\mathcal{M}}_{w_1} \alpha \textit{ there is } w_2 \textit{ such that } R_w w_1 w_2$$
$$\textit{and for every } w_3 \textit{ where } \models^{\mathcal{M}}_{w_3} \alpha \textit{ if } R_w w_2 w_3 \textit{ then } \models^{\mathcal{M}}_{w_3} \beta.$$

So $\alpha \Rightarrow \beta$ is true at w if, essentially, $\alpha \wedge \neg\beta$ is a remoter possibility (according to R_w) than $\alpha \wedge \beta$; equivalently, in considering successively remoter worlds, at some point the only accessible α-worlds are all $\alpha \wedge \beta$-worlds. The underlying accessibility relation (treating R as a binary relation) corresponds to that of S4 [HC68]; [Lam91] presents a closely-related logic. Optionally we could have defined $\Rightarrow$ in term of $\Box$ [Bou94]. The logic has the following characterisation [Bur81]:

Definition 3.2 *The conditional logic $\mathcal{S}$ is the smallest logic with rules of Substitution of (PC) Equivalents and Modus Ponens, and closed under the following axiom schemata*

ID: $\alpha \Rightarrow \alpha$

CC: $(\alpha \Rightarrow \beta \wedge \alpha \Rightarrow \gamma) \implies (\alpha \Rightarrow \beta \wedge \gamma)$

CM: $(\alpha \Rightarrow \beta \wedge \gamma) \implies (\alpha \Rightarrow \beta)$

ASC: $\alpha \Rightarrow \beta \implies (\alpha \Rightarrow \gamma \implies \alpha \wedge \beta \Rightarrow \gamma)$

CA: $(\alpha \Rightarrow \gamma \wedge \beta \Rightarrow \gamma) \implies (\alpha \vee \beta \Rightarrow \gamma)$

Necessity and possibility can be defined in terms of the conditional operator:

Definition 3.3

$$\Box\alpha =_{df} \neg\alpha \Rightarrow \alpha \qquad \textit{and} \qquad \Diamond\alpha =_{df} \neg\Box\neg\alpha.$$

For the remainder of the paper, L will be some logic containing $\mathcal{S}$, and $\mathcal{T}$ will be a consistent theory in this logic. We write $\models_L$ for logical entailment in L, as a reminder that we will be dealing with some logic of a class of logics. So for example 'birds (normally) fly; birds with broken wings (normally) do not fly; penguins are (taxonomically) birds; penguins (normally) do not fly' might be expressed as:

$$B \Rightarrow F, \quad B \wedge BW \Rightarrow \neg F, \quad \Box(P \implies B), \quad P \Rightarrow \neg F.$$

3.3 The Relation of *Relevance*

We can now ask just what it is that relevance is a relation between or among. In the next part we first consider relevance as a relation between arbitrary propositions; we find this characterisation lacking. We subsequently suggest that relevance is rather a relation between a proposition and a conditional.

3.3.1 Relevance as a Relation Between Propositions

Informally, one might claim that α is relevant to γ according to theory $\mathcal{T}$ if $\mathcal{T} \models \alpha \Rightarrow \gamma$. Thus α is relevant to γ if, under conditions determined by α, γ is true. So being a bird is relevant to flight, while being a raven is relevant to being black, and being an albino raven is relevant to being non-black. However this approach is clearly too weak, in that it has nothing to say about compound formulae. A flawed generalisation would be to say that α is relevant to γ if α along with some other information β leads to γ:

α is relevant to γ in $\mathcal{T}$ iff for some β we have that $\mathcal{T} \models \alpha \wedge \beta \Rightarrow \gamma$.

This doesn't work since in the base logic $\alpha \wedge \gamma \Rightarrow \gamma$ is valid, and so this would mean that any formula is relevant to any other formula. We can eliminate the case where β is γ most naturally as follows:

α is relevant to γ in $\mathcal{T}$ iff for some β where $\mathcal{T} \not\models \beta \Rightarrow \gamma$ we have $\mathcal{T} \models \alpha \wedge \beta \Rightarrow \gamma$.

Unfortunately the problem remains: if we now take β to be $\alpha \implies \gamma$ we again have that any formula α is relevant to any formula γ, except possibly in those instances where $\mathcal{T} \models (\alpha \implies \gamma) \Rightarrow \gamma$ holds.

A similar problem arises if we state that α is relevant to γ if α along with some other information β leads to the conclusion γ, whereas $\neg\alpha$ does not:

α is relevant to γ in $\mathcal{T}$ iff

for some β we have $\mathcal{T} \models \alpha \wedge \beta \Rightarrow \gamma$ but $\mathcal{T} \models \neg\alpha \wedge \beta \Rightarrow \neg\gamma$.

Again, if β is taken to be $\alpha \implies \gamma$, it follows that any formula α is relevant to any formula γ. Furthermore, this last proposal isn't even consistent. There is no problem in having theories of the form:

$$\begin{aligned} &\mathcal{T} \models \alpha \wedge \beta \Rightarrow \gamma \text{ and } \mathcal{T} \models \neg\alpha \wedge \beta \Rightarrow \neg\gamma \qquad \text{but} \\ &\mathcal{T} \models \alpha \wedge \delta \Rightarrow \gamma \text{ and } \mathcal{T} \models \neg\alpha \wedge \delta \Rightarrow \gamma. \end{aligned}$$

Thus according to the top two formulas α is relevant to γ while in the bottom two formulas α is not relevant to γ.

However this does suggest a possible solution. As a concrete example, consider the formulae:

$$BW \wedge B \Rightarrow \neg F \quad \text{and} \quad \neg BW \wedge B \Rightarrow F \qquad \text{but} \tag{1}$$

$$BW \wedge NZ \Rightarrow \neg F \quad \text{and} \quad \neg BW \wedge NZ \Rightarrow \neg F. \tag{2}$$

Thus birds with broken wings (BW) normally do not fly, whereas this is not the case if they do not have a broken wing. However New Zealand birds (NZ) normally do not fly, regardless of whether or not they have a broken wing. It seems clear that for flight the 'context' of the antecedent of the conditionals needs to be taken into account – in one case being a bird and in the other being an New Zealand bird. Thus arguably we want to say something like 'having a broken wing is relevant to a bird not flying' whereas 'having a broken wing is not relevant to a New Zealand bird not flying (since typical New Zealand birds don't fly anyway).' We explore this elaboration in the next subsection.

3.3.2 Relevance as a Relation Between a Proposition and a Conditional

The last example suggests that relevance is best interpreted as a relation between a proposition and a conditional, rather than as a relation between two arbitrary propositions. Thus from (1), having a broken wing is relevant to a bird's flying whereas, from (2), being from New Zealand isn't. So the goal now is to specify conditions for α being relevant to $\beta \Rightarrow \gamma$.

Intuitively, α is relevant to a conditional $\beta \Rightarrow \gamma$ if the addition of α changes our belief (confidence, whatever) in γ, given β. We might therefore propose the following as an initial, tentative definition:

Definition 3.4
α is relevant to $\beta \Rightarrow \gamma$ iff: $\mathcal{T} \models \beta \Rightarrow \gamma$ but $\mathcal{T} \models \alpha \wedge \beta \Rightarrow \neg\gamma$.

Definition 3.4 seems to us to constitute a naïve, or perhaps minimal, condition for relevance. In the following sections we will use this tentative definition as a guide against which intuitions and postulates are judged. At the end of Section 4 we present a substantial refinement of this definition.

Even without these refinements, Definition 3.4 alone seems to work relatively well. Consider for example the default theory: birds fly, shorebirds fly, birds with broken wings do not fly, New Zealand birds do not fly, New Zealand shorebirds fly, and New Zealand shorebirds with broken wings do not fly. Symbolically:

$$B \Rightarrow F,\ ShB \Rightarrow F,\ B \wedge BW \Rightarrow \neg F,\ NZ \Rightarrow \neg F,\ NZ \wedge ShB \Rightarrow F,$$
$$NZ \wedge ShB \wedge BW \Rightarrow \neg F.$$

In addition suppose New Zealand birds and shorebirds are necessarily birds, and so $\Box(NZ \implies B)$ and $\Box(ShB \implies B)$. We obtain in this theory that:

ShB	is not relevant to	$B \Rightarrow F$.
ShB	is relevant to	$NZ \Rightarrow \neg F$.
ShB	is not relevant to	$B \wedge BW \Rightarrow \neg F$.

In addition, returning to the ravens-being-black example, we have:

Al is relevant to $R \Rightarrow Bl$ but $\neg Al$ is not relevant to $R \Rightarrow Bl$.

So, our naïve notion of relevance seems to be heading in the right direction. However, one difficulty is that we have no standards against which to judge this definition. We could, for example, continue to examine specific examples to determine the suitability of the definition. But then we would just have sets of examples that bolster the acceptability of the definition; one might justifiably be called upon to justify the examples as constituting a broad and appropriate set. Instead, we ask what *principles* ought any definition of relevance obey. We can then consider how a given principle interacts with others, or what consequences follow from the set of principles. Consequently, in the next section we develop a set of postulates that govern relevance, using the tentative definition as a point of departure. We then reformulate the tentative definition so that it is in accord with these principles.

4 Principles of Relevance

This section proposes a number of postulates governing relevance with respect to subjunctive conditionals, or, in our framework, 'relevance-within-a-theory-expressed-in-a-conditional-logic.' In the first subsection we examine basic postulates concerned with those things that are relevant-given-certain-logical-truths, or relevant-given-some-other-relevant-condition. In the second subsection we consider how relevant conditions may be combined to yield further relevant conditions. Finally we present explicit definitions that satisfy these conceptions of relevance.

4.1 Basic Principles

'Relevant' and 'irrelevant' will be taken as mutually exclusive and jointly exhaustive notions:

P0: (α is relevant to $\beta \Rightarrow \gamma$) iff (α is not irrelevant to $\beta \Rightarrow \gamma$).

So if we can take care of 'relevant' we will also take care of 'irrelevant'.

We have argued that relevance is a semantic notion, and so we require that any formalisation of relevance not depend on the syntactic form of formulae. In our view, theories relying on the syntactic form of sentences ([SG87], for example) are not really talking about relevance but instead about some relation between certain syntactic objects – such as being known by a user to be interesting for one reason or another. We believe that relevance is a relation that holds between properties and (default) relationships, both of which are items in reality – or at least, in the reality presumed by some (default) theory. At best, a syntactic account could capture only a portion of this aspect of reality, namely that portion which coincided with a purely semantic account.[5] For, in any respect that the syntactic account deviated from the semantic account, the former would not be describing the reality of relevance relationships, but rather an agent's understanding or beliefs about how this relationship can be described by sentences.

[5] At worst it merely describes certain properties of inscriptions, such as being written in red or in green ink – and this without even any logical principles like 'each inscription can only be written in one ink colour' (much less with principles such as 'logically equivalent sentences are written in the same colour'). The relation of relevance is *not* like being written in red ink. It is a relation that certain properties bear to certain (default) relationships.

In a purely semantic account, when two terms provably designate the same property – that is, provably there is but one property – then whatever is true about this property, is true about it regardless of the manner of designation. Consequently we require:

Irrelevance of Syntax

If α is relevant to $\beta \Rightarrow \gamma$ and

$$\mathcal{T} \models_L \Box(\alpha \equiv \delta) \text{ and } \mathcal{T} \models_L \Box(\beta \equiv \eta) \text{ and } \mathcal{T} \models_L \Box(\gamma \equiv \zeta)$$

then δ is relevant to $\eta \Rightarrow \zeta$.

Since any conditional logic that we consider allows substitutivity of $\Rightarrow$-equivalents in the antecedent of a conditional,[6] we can replace the Irrelevance of Syntax Principle with the following stronger version.

P1: If α is relevant to $\beta \Rightarrow \gamma$ and

$$\mathcal{T} \models_L \Box(\alpha \equiv \delta) \text{ and } \mathcal{T} \models_L \beta \Leftrightarrow \eta \text{ and } \mathcal{T} \models_L \Box(\gamma \equiv \zeta)$$

then δ is relevant to $\eta \Rightarrow \zeta$.

An instance of **P1** is the following:

If α is relevant to $\beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \Box(\gamma \equiv \zeta)$ then α is relevant to $\beta \Rightarrow \zeta$.

This cannot in general be strengthened to:

If α is relevant to $\beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \Box(\gamma \implies \zeta)$ then α is relevant to $\beta \Rightarrow \zeta$.

To see that this is unacceptable, consider the fact that male cardinals are birds that are uniformly coloured red. Imagine that some disease changes the coloration to yellow; hence this disease is relevant to male cardinals being red. However, necessarily red things are coloured, but we would not want to say that the disease is relevant to cardinals being coloured (since they are coloured whether or not they have the disease).

On the other hand, no weakening of the antecedent of a conditional should be relevant to the conditional. The strongest assertion of this principle is the following:

P2: If $\mathcal{T} \models_L \beta \Rightarrow \alpha$ then α is not relevant to $\beta \Rightarrow \gamma$.

Since we have the meta-theorem

$\mathcal{T} \models_L \Box(\beta \implies \alpha)$ infer $\mathcal{T} \models_L \beta \Rightarrow \alpha$,

it follows immediately that:

$$\text{If } \mathcal{T} \models_L \Box(\beta \implies \alpha) \text{ then } \alpha \text{ is not relevant to } \beta \Rightarrow \gamma. \tag{3}$$

If $\top$ is a designated propositional sentence true at all worlds in all models, then taking α as $\top$ in (3) we immediately obtain the result that:

$$\top \text{ is not relevant to } \beta \Rightarrow \gamma. \tag{4}$$

If α and β are mutually contradictory, then α has nothing to say about whether $\beta \Rightarrow \gamma$:

[6] That is, $(\alpha \Rightarrow \beta \wedge \beta \Rightarrow \alpha) \implies (\alpha \Rightarrow \gamma \equiv \beta \Rightarrow \gamma)$ is valid in these logics.

P3: If $\mathcal{T} \models_L \Box\neg(\alpha \wedge \beta)$ then α is not relevant to $\beta \Rightarrow \gamma$.

We obtain as corollaries of this principle that:

$\neg\beta$ is not relevant to $\beta \Rightarrow \gamma$.

$\neg\top$ is not relevant to $\beta \Rightarrow \gamma$.

If $\mathcal{T} \models_L \Box\neg\beta$ then α is not relevant to $\beta \Rightarrow \gamma$.

We also believe that any definition of relevance must assert that every default conditional has some properties that are irrelevant to it:

P4: If $\mathcal{T} \models_L \Box(\alpha_1 \vee \cdots \vee \alpha_n)$ then for a conditional $\beta \Rightarrow \gamma$

there is a k, $1 \leq k \leq n$, such that α_k is not relevant to $\beta \Rightarrow \gamma$.

That is, if we know that $\alpha_1, \ldots, \alpha_n$ essentially cover all possibilities, then for any given conditional, some α_k is irrelevant to that conditional. Since $\mathcal{T} \models_L \Box(\alpha \vee \neg\alpha)$ we obtain:

If α is relevant to $\beta \Rightarrow \gamma$ then $\neg\alpha$ is not relevant to $\beta \Rightarrow \gamma$.

Being albino is relevant to a raven being black, since albinoism leads us to change our opinion concerning a raven's coloration, and so by **P4** non-albinoism is irrelevant to it. This accords with intuition, that lack of albinoism does not lead us to change our opinion about a raven's coloration. Similarly for my visit to the dentist's office: given that I have some plan formulated to get to the office, the fact that the elevator is out of order is relevant to my plan, and so therefore the fact that it isn't is not.[7] Note that letting $n = 1$ in **P4**, yields a condition equivalent to (4).

Lastly, note that this postulate is in accord with our tentative definition, Definition 3.4, since the set of sentences

$$\{\Box(\alpha_1 \vee \alpha_2),\ \ \Diamond\beta,\ \ \beta \Rightarrow \gamma,\ \ \alpha_1 \wedge \beta \Rightarrow \neg\gamma,\ \ \alpha_2 \wedge \beta \Rightarrow \neg\gamma\}$$

is unsatisfiable. Hence, if we were to accept the tentative definition, we could not consistently deny **P4**.

The preceding postulates *must* hold, given our understanding of 'relevance'; if they don't, one does not have an adequate account of (say) 'causal relevance.' The next postulate is one which may or may not be adopted, according to one's other intuitions:

P5: If α is relevant to $\beta \Rightarrow \gamma$ then α is relevant to $\beta \Rightarrow \neg\gamma$.

That is, in accepting the postulate, one would say that α is relevant to *whether or not* γ holds in the case of β. Alternatively, in not accepting the postulate, one would say only that α is relevant to *the fact that* γ holds in the case of β. Thus in the first case, one would take the position that albinoism is relevant to a raven's being not black, since it is certainly relevant to a raven's being black. In the second case one would simply say that albinoism is

[7]This last relation is in contrast with the Gärdenfors postulate, where p is relevant to r on evidence e iff $\neg p$ is relevant to r. See Section 5 for a discussion.

relevant to the assertion that a raven is black. These senses of relevance are referred to as 'whether-relevance' and 'that-relevance'.

Note that this postulate is inconsistent with the naïve definition. For if we have that α is relevant to $\beta \Rightarrow \gamma$ then by **P5** we would also have that α is relevant to $\beta \Rightarrow \neg\gamma$. But then by the naïve definition we would obtain $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$ from which we obtain $\mathcal{T} \models_L \Box\neg(\alpha \wedge \beta)$, contradicting **P3**. In our 'official' definition of relevance we do not adopt **P5**, although we do provide an explicit definition for a notion of relevance incorporating it.

Consider next a principle of augmentation of relevant propositions.

AUG: If α is relevant to $\beta \Rightarrow \gamma$ then $\alpha \wedge \delta$ is relevant to $\beta \Rightarrow \gamma$.

The difficulty with this principle is that it makes too many things relevant. For example, if all we know about birds is that birds normally fly, but birds with broken wings do not, then it seems that the *only* thing relevant to birds flying is having a broken wing. Having a broken wing and being green is not relevant; nor is having a broken wing and having stayed up all night; nor is having a broken wing together with the current state of the Tokyo stock market. Clearly, if we were reasoning about whether a bird might fly or not, we would want to consider only whether it had a broken wing or not, and not this information conjoined with irrelevant facts. Hence we require that a relevant condition be *parsimonious*, in that it not contain an irrelevant condition as a part. This is precisely the stance taken in diagnostic reasoning in Artificial Intelligence for instance, where one is interested in a *minimal* diagnosis. In this realm, having influenza could be a perfectly adequate diagnosis accounting for symptoms of fever, muscular pain, etc. However having influenza and a lawn that needs cutting, while equally well accounting for the symptoms, nonetheless does not constitute an adequate diagnosis. We will call a definition of relevance that accepts **AUG** *promiscuous* relevance.

Note further that **AUG** does not sit well with our naïve definition, Definition 3.4, according to which it ought to be possible for having a broken wing to be relevant to a bird flying, but for having a broken wing and being green to not be relevant to a bird flying. But **AUG** requires that any such strengthening of a relevant condition must be relevant. Many researchers find conditional logics to be useful precisely because they allow a conditional $\beta \Rightarrow \gamma$ to be true while $\alpha \wedge \beta \Rightarrow \gamma$ is not. But **AUG**, in concert with Definition 3.4, will not allow this: if α is relevant to $\beta \Rightarrow \gamma$, then $\alpha \wedge \beta \Rightarrow \neg\gamma$ by Definition 3.4. But by **AUG**, $\alpha \wedge \delta$ is also relevant to $\beta \Rightarrow \gamma$, for any δ; and so by Definition 3.4, $\alpha \wedge \delta \wedge \beta \Rightarrow \neg\gamma$. And all further conjuncts added to $\alpha \wedge \beta$ in the antecedent will yield $\neg\gamma$ in the consequent. None of the characteristic shifting back and forth can happen. No further strengthening of α can change the consequent of this conditional, and thus much of the interest of the conditional logic framework is lost.

It might be argued that **AUG** is a reasonable relationship, since clearly birds fly whereas a bird that is green and has a broken wing does not. However *this* argument seems to confuse the assertion

$$BW \wedge Gr \quad \text{is relevant to} \quad B \Rightarrow F$$

with the assertion

$$BW \quad \text{is relevant to} \quad B \wedge Gr \Rightarrow F.$$

This last assertion is clearly acceptable, since green birds normally fly whereas green birds with broken wings normally do not. (This is despite the presumed fact that greenness is irrelevant to a bird flying.)

4.2 Combining Relevant Properties

An issue that we have not discussed is that of strengthenings or weakenings of relevant conditions, and it is to this issue that we now turn. Consider strengthenings of relevant conditions first. Recall that according to the tentative definition,

$$ShB \wedge BW \text{ is not relevant to } NZ \Rightarrow \neg F \tag{5}$$

because, while New Zealand birds don't fly, neither do New Zealand shorebirds with broken wings. However, if we consider the individual conjuncts ShB and BW, the tentative definition says

$$ShB \quad \text{is relevant to} \quad NZ \Rightarrow \neg F \text{ and} \tag{6}$$
$$BW \quad \text{is relevant to} \quad NZ \wedge ShB \Rightarrow F \tag{7}$$

(since New Zealand shorebirds do fly, but New Zealand shorebirds with broken wings don't fly). Arguably then $ShB \wedge BW$ should be relevant to $NZ \Rightarrow \neg F$ since, given NZ along with $ShB \wedge BW$, we would conclude $\neg F$ by default, but for different reasons than if we were just given NZ alone. Thus $ShB \wedge BW$ *should* be relevant to $NZ \Rightarrow \neg F$, since ShB is.

On the other hand, intuitively we have that

$$BW \quad \text{is not relevant to} \quad NZ \Rightarrow \neg F \text{ and}$$
$$ShB \quad \text{is not relevant to} \quad NZ \wedge BW \Rightarrow F.$$

Thus we deal with the conjunction of individually relevant properties by breaking them up in a particular order, but not necessarily in some other order. This in turn suggests that we, in part, will ultimately deal with a compositional definition of relevance, where more complex conditions are built from simpler ones. To capture this we add a weakened version of **AUG**:

P6: If α is relevant to $\beta \Rightarrow \gamma$ and δ is relevant to $\alpha \wedge \beta \Rightarrow \neg\gamma$ then $\alpha \wedge \delta$ is relevant to $\beta \Rightarrow \gamma$.

However we reject the following variant of **P6**:

$$\text{If } \alpha \text{ and } \delta \text{ are individually relevant to } \beta \Rightarrow \gamma \text{ then } \alpha \wedge \delta \text{ is relevant to } \beta \Rightarrow \gamma. \tag{8}$$

While having a broken wing is relevant to a bird flying, and having feet set in concrete is also, we do not thereby admit that having a broken wing *and* feet set in concrete constitutes a relevant condition for a bird flying. Rather, this last condition combines two distinct conditions.[8] As remarked above, a similar stance is taken in diagnostic reasoning. For example

[8]This means that we have some notion as to what constitutes a 'distinct condition'. For example, if the theory also had an atomic letter that (intuitively) meant $BW \wedge FC$, then we might want to admit this as a condition. We return to this point shortly.

having influenza might be a diagnosis accounting for a blocked nose, itchy eyes, etc., and having allergies might account for precisely the same set of symptoms. However having influenza-and-allergies, while undoubtedly accounting for the symptoms, would not constitute a distinct diagnosis, but would be considered a conflation of two individual diagnoses. Gärdenfors [Gär78, p. 358] rejects a similar postulate in the context of his probabilistic account of relevance, arguing that it leads to unintuitive results.

There are two further plausible-looking principles, intended to deal with combinations of relevant conditions, that we consider here. First:

P7: If α is relevant to $\beta \Rightarrow \gamma$ and δ is relevant to $\beta \Rightarrow \gamma$ then $\alpha \vee \delta$ is relevant to $\beta \Rightarrow \gamma$.

Consider a default theory that asserts among other things that birds from the Antarctic don't fly, nor do New Zealand birds:

$$\{B \Rightarrow F,\ \ B \wedge Ant \Rightarrow \neg F,\ \ B \wedge NZ \Rightarrow \neg F,\ \ B \wedge (Ant \vee NZ) \Rightarrow \neg F\} \tag{9}$$

(The last formula is in fact a logical consequence of the previous two.) Intuitively, being an Antarctic bird is relevant to a bird flying, as is being a New Zealand bird. By **P7** being an Antarctic-or-New-Zealand bird would also be relevant. This principle is supported by Definition 3.4, according to which all of Ant, NZ, $(Ant \vee NZ)$ would be relevant to $B \Rightarrow F$.

However, we reject this principle for the same reasons that we rejected (8), that something like $(Ant \vee NZ)$ isn't a relevant property but rather is *composed* of relevant properties. If we were to accept **P7**, then for a conditional such as $B \Rightarrow F$, if we had n 'basic' relevant properties, such as being from New Zealand, having a broken wing, having feet set in concrete, etc., then this principle would add on the order of 2^n other relevant properties, and this intuitively seems unreasonable. So this is a prima facie reason to reject **P7**. On the other hand, a rejection of **P7** means that Definition 3.4 is too strong, since in the logics that we deal with $((\alpha \Rightarrow \gamma) \wedge (\beta \Rightarrow \gamma)) \implies (\alpha \vee \beta \Rightarrow \gamma)$ is valid. So if we reject **P7** then we must weaken our tentative definition.

However let us leave this principle for the moment and consider another closely related one:

P8: If α is relevant to $\beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \Box(\alpha_1 \vee \cdots \vee \alpha_n)$ then:

if for every k, $1 \leq k \leq n$, we have that $\alpha \wedge \alpha_k$ is relevant to $\beta \Rightarrow \gamma$,
then, for some k, $1 \leq k \leq n$ we have that $\mathcal{T} \models_L (\alpha \wedge \beta) \Leftrightarrow (\alpha \wedge \beta \wedge \alpha_k)$.

In **P8** for $n = 2$ we obtain:

If α is relevant to $\beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \Box(\alpha_1 \vee \alpha_2)$ then

if $\alpha \wedge \alpha_1$ is relevant to $\beta \Rightarrow \gamma$ and $\alpha \wedge \alpha_2$ is relevant to $\beta \Rightarrow \gamma$

$$\text{then } \mathcal{T} \models_L (\alpha \wedge \beta) \Leftrightarrow (\alpha \wedge \beta \wedge \alpha_1) \text{ or } \mathcal{T} \models_L (\alpha \wedge \beta) \Leftrightarrow (\alpha \wedge \beta \wedge \alpha_2). \tag{10}$$

P8 is actually quite straightforward, although it appears somewhat complex. Consider the version given in (10). This says that if α is relevant to $\beta \Rightarrow \gamma$ and $\alpha \wedge \alpha_1$ and $\alpha \wedge \alpha_2$ are also, then in fact in the context of β we have that $\alpha \wedge \alpha_1$ or $\alpha \wedge \alpha_2$ is conditionally equivalent to

α. An example is given by (6) and (7). We have that ShB is relevant to $NZ \Rightarrow \neg F$ and (by virtue of **P6**) $BW \wedge ShB$ is relevant to $NZ \Rightarrow \neg F$. We also have $\neg BW \wedge ShB$ is relevant to $NZ \Rightarrow \neg F$, but only because here $\neg BW \wedge ShB \wedge NZ \Leftrightarrow ShB \wedge NZ$ is true. So in this case when we say that $\neg BW \wedge ShB$ is relevant to $NZ \Rightarrow \neg F$, this is in fact the same assertion as ShB is relevant to $NZ \Rightarrow \neg F$.

So assume we have that α is relevant to $\beta \Rightarrow \gamma$ but $\mathcal{T} \not\models_L (\alpha \wedge \beta) \Leftrightarrow (\alpha \wedge \beta \wedge \alpha_1)$ and $\mathcal{T} \not\models_L (\alpha \wedge \beta) \Leftrightarrow (\alpha \wedge \beta \wedge \alpha_2)$. Then an instance of (10) is

If α is relevant to $\beta \Rightarrow \gamma$ then

$$\text{if } \alpha \wedge \alpha_1 \text{ is relevant to } \beta \Rightarrow \gamma \text{ then } \alpha \wedge \neg\alpha_1 \text{ is not relevant to } \beta \Rightarrow \gamma. \tag{11}$$

Here we have that, if it were claimed that $\alpha \wedge \alpha_1$ was relevant to $\beta \Rightarrow \gamma$ and $\alpha \wedge \neg\alpha_1$ was also, then it would seem that α_1 plays no role in determining whether γ holds, given β. Thus, under a principle of parsimony, we would hold instead that α alone was relevant to $\beta \Rightarrow \gamma$. Consider a default theory that asserts, among other things, that birds fly; birds that have a broken wing and are green don't fly; birds that have a broken wing and are nongreen don't fly; and birds with a broken wing don't fly:

$$\{B \Rightarrow F,\ \ B \wedge BW \wedge Gr \Rightarrow \neg F,\ \ B \wedge BW \wedge \neg Gr \Rightarrow \neg F,\ \ B \wedge BW \Rightarrow \neg F\} \tag{12}$$

(The last formula is again a logical consequence of the previous two.) Intuitively, the only thing relevant to a bird flying is having a broken wing. So we would want to have

$$\begin{array}{lll} BW \wedge Gr & \text{is not relevant to} & B \Rightarrow F \text{ and} \\ BW \wedge \neg Gr & \text{is not relevant to} & B \Rightarrow F. \end{array} \tag{13}$$

Unfortunately examples (9) and (12) are essentially the same – one need just substitute $BW \wedge Gr$ for Ant and $BW \wedge Gr$ for NZ to transform the former into the latter. Yet the desired relevant conditions differ in these examples. So depending on the example, one has conflicting intuitions. More formally, consider the contrapositive of (11):

If $\alpha \wedge \alpha_1$ is relevant to $\beta \Rightarrow \gamma$ and $\alpha \wedge \neg\alpha_1$ is relevant to $\beta \Rightarrow \gamma$ then

$$\alpha \text{ is not relevant to } \beta \Rightarrow \gamma. \tag{14}$$

This last relation is in direct conflict with **P7**, when one substitutes $(\alpha \wedge \alpha_1)$ for α and $(\alpha \wedge \neg\alpha_1)$ for δ in **P7**, and then reduces to an equivalent form. So **P7** and (14) together entail that

$$\text{If } \alpha \wedge \alpha_1 \text{ is relevant to } \beta \Rightarrow \gamma \text{ then } \alpha \wedge \neg\alpha_1 \text{ is not relevant to } \beta \Rightarrow \gamma. \tag{15}$$

However, this (derived) principle leads to unacceptable consequences. Consider again Example (9). Clearly we would want that

$$\begin{array}{lll} Ant & \text{is relevant to} & B \Rightarrow F \text{ and} \\ NZ & \text{is relevant to} & B \Rightarrow F \end{array} \tag{16}$$

A substitution instance of (15) (substituting $(Ant \vee NZ)$ for α, and Ant for α_1) yields:

If $(Ant \vee NZ) \wedge Ant$ is relevant to $B \Rightarrow F$
then $(Ant \vee NZ) \wedge \neg Ant$ is not relevant to $B \Rightarrow F$

and so:

$$\text{If } Ant \text{ is relevant to } B \Rightarrow F \text{ then } NZ \text{ is not relevant to } B \Rightarrow F. \tag{17}$$

(Ant in the *if* clause of (17) is logically equivalent to the corresponding relevance formula in the preceding equation, while NZ in the *then* clause of (17) is equivalent to its corresponding relevance formula because the background theory contains $\Box(NZ \implies \neg Ant)$. (17) of course is inconsistent with (16).)

Thus we cannot hold both **P7** and **P8**; but since these two principles are not exhaustive, we have the option of rejecting either principle or both. As indicated earlier, we choose to reject **P7**, partly because it leads to a multiplicity of relevant conditions, and partly because it cannot be held along with **P8** – and presumably we want **P8** to take care of (13). Furthermore, if we reject **P8** we end up with a notion of relevance that is overly weak, in that we cannot now deal with (13). Consequently our postulate set consists of **P0** – **P4**, optionally **P5**, **P6**, and **P8**.

4.3 An Explicit Definition of Relevance

This section presents an explicit definition (or more accurately, definitions) of relevance that satisfies our adopted set of postulates of the preceding subsections. Our initial tentative definition, Definition 3.4, provides a base case for relevance, except that it does not address our requirement of parsimony. Consider again (12). Clearly BW is relevant to $B \Rightarrow F$, and arguably $BW \wedge Gr$ is not. BW is relevant, since it would lead us to change our mind concerning whether a particular bird flies, but the addition of Gr would not, either about birds flying or about broken-winged birds not flying. So by parsimony we would like to say that $BW \wedge Gr$ is not relevant to $B \Rightarrow F$.

Unfortunately this presents us with a problem. Intuitively we want to say that α is relevant to $\beta \Rightarrow \gamma$ if α is the 'least strong' condition such that $\alpha \wedge \beta \Rightarrow \neg\gamma$ is true. So this would let us say that having a broken wing is relevant, while having-a-broken-wing-and-being-green is not. The difficulty is that if we accept such a condition, then in (9) we would be directed to conclude that neither NZ nor Ant are relevant conditions, but $Ant \vee NZ$ is, since $Ant \vee NZ$ is weaker. This is clearly not acceptable.

One solution is to note that the natural representation of (9) and (12) is, respectively:

$$\{B \Rightarrow F,\;\; B \wedge Ant \Rightarrow \neg F,\;\; B \wedge NZ \Rightarrow \neg F\}$$

$$\{B \Rightarrow F,\;\; B \wedge BW \Rightarrow \neg F\}$$

If we deal with just these sentences, then it is straightforward to specify the appropriate relevant conditions. The difficulty with this solution of course is that it requires a wholesale embracing of a syntactic approach. Our goal at the outset was to provide a semantic characterization. So we reject this possibility.

In the solution we propose below, we retain (more or less) a semantic characterization; however, we obtain a characterization of relevance that is dependent on the *language* expressing the theory. Essentially our solution is to restrict the set of propositions from which relevant properties may be drawn, and in this fashion exclude the anomalous relevance conditions of (9).

Informally, the difference between (9) and (12) is traceable to the difference between the propositions denoted by B, $B \wedge Gr$ and Ant on the one hand, and by $Ant \vee NZ$ on the other. The former would seem to denote 'reasonable' coherent properties, whereas the latter would seem to not denote any specific or coherent property. So the disjunction of $BW \wedge Gr$ and $BW \wedge \neg Gr$ (which of course is just BW) would seem to denote a coherent property, while the disjunction of Ant and NZ does not.[9] This suggests allowing only propositions of the first type as candidates for relevant properties.

We make this more formal by first identifying the *basic propositions* of a theory:

Definition 4.1 *A sentence α denotes a* basic proposition *in theory $\mathcal{T}$ iff for some β, where β is a literal or a conjunction of literals, we have that $\mathcal{T} \models_L \Box(\alpha \equiv \beta)$.*

Thus in a theory wherein animals don't walk upright, but humans do, and humans are necessarily male or female:

$$An \Rightarrow \neg Up, \quad An \wedge Hu \Rightarrow Up, \quad \Box(Hu \equiv (MP \vee FP))$$

basic propositions would include An, Hu, Up, MP, FP, and their negations as well as $(MP \vee FP)$. Relevance can now be restricted to this set of basic propositions; we express this as an additional postulate, introduced to facilitate the development of an explicit definition.

PL0: If α is not a basic proposition in theory $\mathcal{T}$ then α is not relevant to $\beta \Rightarrow \gamma$.

According to **PL0**, being a basic proposition is a necessary but not sufficient condition for relevance. So the set of basic propositions provides us with a 'pool' of candidates from which relevant properties can be drawn. $Ant \vee NZ$ is excluded as a (potential) relevant condition to $B \Rightarrow F$ because it isn't a basic proposition. However Ant and NZ can be individually relevant. BW, $BW \wedge Gr$ and $BW \wedge \neg Gr$ are all basic propositions, but only BW would be relevant to $B \Rightarrow F$. In addition one could, if it were deemed desirable, allow a restricted version of (**P7**) by accepting $\alpha_1 \vee \alpha_2$ as relevant if α_1 and α_2 designated basic propositions.

In accepting **PL0** we lose language-independence, in that relevance conditions depend on our choice of language, and specifically on our choice of primitive propositions. While this isn't ideal, it does seem necessary to have (9) and (12) come out right. So while our account is now in part syntactic, our definition will still be phrased in terms of the set of logical consequences of a theory.

We have the following definitions. Irrelevance is defined in terms of relevance; relevance is defined in part in terms of a preliminary definition of minimal relevance.

[9] This distinction between 'coherent' and 'incoherent' properties is reminiscent of that between 'projectible' and 'non-projectible' properties in the philosophy of science [Goo73]. So the argument might be put forward that a language for a theory comes with certain inductive commitments given in the choice of atomic sentences – that these atomic sentences denote coherent properties. Conjunctions of such coherent properties would also seem to denote coherent properties (or would be empty) whereas disjunctions would not necessarily.

Definition 4.2 *α is irrelevant to $\beta \Rightarrow \gamma$ in theory $\mathcal{T}$ iff α is not relevant to $\beta \Rightarrow \gamma$.*

Definition 4.3 *α is* minimally relevant *to $\beta \Rightarrow \gamma$ in theory $\mathcal{T}$ iff*

1. *α is a basic proposition in $\mathcal{T}$,*
2. *$\mathcal{T} \models_L \beta \Rightarrow \gamma$, $\mathcal{T} \not\models_L \Box\neg(\alpha \wedge \beta)$ and $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$, and*
3. *If α' is a basic proposition where $\mathcal{T} \models_L \Box(\alpha \implies \alpha')$ but $\mathcal{T} \not\models_L \Box(\alpha \equiv \alpha')$ then $\mathcal{T} \not\models_L \alpha' \wedge \beta \Rightarrow \neg\gamma$.*

So if α is minimally relevant to $\beta \Rightarrow \gamma$ then for no entailed but nonequivalent basic proposition α' is it true that $\mathcal{T} \models \alpha' \wedge \beta \Rightarrow \neg\gamma$.

Definition 4.4 *The set of propositions* relevant *to $\beta \Rightarrow \gamma$ in theory $\mathcal{T}$ is the least set satisfying:*

α is relevant *to $\beta \Rightarrow \gamma$ in theory $\mathcal{T}$ iff*

1. *α is minimally relevant to $\beta \Rightarrow \gamma$ in theory $\mathcal{T}$ or*
2. *there is α' such that $\mathcal{T} \models \Box(\alpha \implies \alpha')$ where α' is relevant to $\beta \Rightarrow \gamma$ and α is relevant to $\alpha' \wedge \beta \Rightarrow \neg\gamma$.*

The notion of relevance expressed in Definition 4.4, part 2. allows chains of successively stronger propositions, each directly relevant to the next in the 'context' of β. Thus, NZ is relevant to $B \Rightarrow F$, and so $NZ \wedge ShB$ is relevant to $B \Rightarrow F$ even though both $B \Rightarrow F$ and $NZ \wedge ShB \wedge B \Rightarrow F$ are true. We obtain:

Theorem 4.1 *Definition 4.4 satisfies* **P0 – P4, P6, P8, PL0**.

Definition 4.4 is what we referred to as *that-relevance* in discussing **P5**. A definition of *whether-relevance* is easily obtained from Definitions 4.3 and 4.4:

Definition 4.5 *α is* minimally whether-relevant *to $\beta \Rightarrow \gamma$ iff:*

1. *α is minimally relevant to $\beta \Rightarrow \gamma$ or*
2. *α is minimally relevant to $\beta \Rightarrow \neg\gamma$.*

Definition 4.6 *α is* whether-relevant *to $\beta \Rightarrow \gamma$ iff*

1. *α is minimally whether-relevant to $\beta \Rightarrow \gamma$ in theory $\mathcal{T}$ or*
2. *there is α' such that $\mathcal{T} \models \Box(\alpha \implies \alpha')$ where α' is whether-relevant to $\beta \Rightarrow \gamma$ and α is whether-relevant to $\alpha' \wedge \beta \Rightarrow \neg\gamma$.*

Theorem 4.2 *Definition 4.6 satisfies* **P1 – P6, P8, PL0**.

Lastly we have:

Theorem 4.3 *Relevance (whether-relevance) is* non-trivial*; that is, there are theories with extant relevance (whether-relevance) conditions.*

5 Discussion

Arguably Definition 4.4 captures a reasonable notion of 'relevant', with Definition 4.6 providing a variant. In addition to previously-mentioned relations, we also obtain the following, in the obvious theory:

BW	is relevant to	$B \Rightarrow F$
NZ	is relevant to	$B \Rightarrow F$
$NZ \wedge BW$	is not relevant to	$B \Rightarrow F$
$BW \wedge Gr$	is not relevant to	$B \Rightarrow F$
BW	is relevant to	$B \wedge Gr \Rightarrow F$
BW	is not relevant to	$NZ \Rightarrow \neg F$
ShB	is relevant to	$NZ \Rightarrow \neg F$
$ShB \wedge BW$	is relevant to	$NZ \Rightarrow \neg F$
$BW \vee NZ$	is not relevant to	$B \Rightarrow F$

This approach can be contrasted with quantitative approaches. In classical probability theory, a proposition p is conditionally independent of r given e, written $I(p, e, r)$, if

$$P(r \mid p \wedge e) = P(r \mid e) \text{ whenever } P(p \wedge e) \neq 0.$$

[Pea88] provides an axiomatic characterisation of independence:

Symmetry: $I(p, e, r) \equiv I(r, e, p)$.

Decomposition: $I(p, e, r \wedge s) \implies I(p, e, r) \wedge I(p, e, s)$.

Weak Union: $I(p, e, r \wedge s) \implies I(p, e \wedge s, r)$.

Contraction: $I(p, e, r) \wedge I(p, e \wedge r, s) \implies I(p, e, r \wedge s)$.

If we were to equate conditional independence with irrelevance in our approach then $I(p, e, r)$ would be represented as 'p is not relevant to $e \Rightarrow r$'. And in this case we would want none of these axioms in our approach. This is not too surprising since conditional independence is simply too weak to capture any reasonable notion of relevance. The **Symmetry** axiom in our approach would lead to our adopting 'being a bird is relevant to a green thing flying' iff 'flying is relevant to a green thing being a bird'. If it were the case that most green flying things were insects, then clearly we would want to accept the first while not accepting the second. Similar examples are easily constructed for the other axioms. Moreover, **Decomposition** subsumes **AUG** in the presence of **Symmetry**.

Peter Gärdenfors [Gär78] takes conditional independence as a point of departure for a probability-based approach to relevance. He provides a definition for 'p is relevant on evidence e to r'; in our notation this would be expressed as 'p is relevant to $e \Rightarrow r$'. Concerning postulates of relevance, there are a number of points in common, as well as several differences, between his approach and ours. To begin with, both approaches take relevance and irrelevance as being mutually exclusive and exhaustive. As well, propositions that are necessarily true or necessarily false are irrelevant in our approach; tautologies and contradictions are irrelevant in Gärdenfors'.

With respect to differences, as we mentioned earlier, we don't adopt anything like his postulate:

R2: p is relevant on evidence e to r iff $\neg p$ is relevant on evidence e to r.

Rather we have two contrasting principles, the first given as a corollary to **P4**:

> If δ is relevant to $\beta \Rightarrow \gamma$ then $\neg\delta$ is not relevant to $\beta \Rightarrow \gamma$.

Secondly, we have **P5**, which we left open as to whether it should be adopted or not:

> If α is relevant to $\beta \Rightarrow \gamma$ then α is relevant to $\beta \Rightarrow \neg\gamma$.

Also, Gärdenfors has:

R7: If p is irrelevant on evidence e to r, and q is irrelevant on evidence e to r, then $p \wedge q$ is irrelevant on evidence e to r,

or equivalently:

> If $p \wedge q$ is relevant on evidence e to r then p is relevant on evidence e to r or q is relevant on evidence e to r,

In our approach we can represent a counterexample such as: To get ahead in this world (W) a person (P) has to be both hard-working (H) and intelligent (I); neither one alone is enough. The natural representation of this truism is

> I is not relevant to $P \Rightarrow \neg W$
>
> H is not relevant to $P \Rightarrow \neg W$
>
> $H \wedge I$ is relevant to $P \Rightarrow \neg W$.

So the two jointly are relevant, yet neither one is individually, contrary to Gärdenfors's **R7**.

Gärdenfors ends his paper with the open question of whether

> Given background evidence e, p is relevant to r iff r is relevant to p

is true. In our approach we would represent this query as whether:

> p is relevant to $e \Rightarrow r$ iff r is relevant to $e \Rightarrow p$

to which we would answer 'no'. Being an elephant (E) is relevant to a living thing's (L) not being a mammal (M), because $L \Rightarrow \neg M$ and $L \wedge E \Rightarrow M$ are true. On the other hand, both $L \Rightarrow \neg E$ and $L \wedge M \Rightarrow \neg E$ are true, and hence M, being a mammal, is *not* relevant to $L \Rightarrow \neg E$, a living thing's not being an elephant.

The conclusion to be drawn from all of this is that the probabilistic and conditional-based approaches to relevance, while employing a common term 'relevant', nonetheless are discussing quite different things. Our approach uses a sense that can be called 'causal' relevance and is appropriate in the context of subjunctive conditionals. Gärdenfors, along with other probabilistic or numeric-based approaches, uses a sense of 'relevant' which can be called 'predictive' relevance. In these numeric-based approaches, both albinoism and non-albinoism (for example) are relevant to a raven being black, since knowing either property would alter one's certainty concerning the raven's predicted colour. In our account, albinoism is relevant to a raven's being black since, in some fashion or another, it constitutes a *cause* of a raven being non-black. Non-albino is not relevant, since it has no causal effect on a raven's normal colour.

Arguably, the concept of relevance developed here is more broadly applicable that that found in prior accounts. The examples concerning when it may be permissible to exceed the speed limit and how I might get to my dentist's office clearly involve a notion of relevance, but in a setting where probability theory plays no role.

Goodman's notion of *relative aboutness* [Goo61], which might be thought to be comparable to our notion of relevance, has quite different properties from ours. For example, relative aboutness is symmetric, and so one would state for example that 'S and T are *about* k', relative to each other. But if $\models_L S \implies T$, then S and T are not *about* anything relative to each other. Relevance, however, does allow such a relationship; for example both NZ and $NZ \land ShB$ are relevant to $B \Rightarrow F$. So here also, Goodman's *relative aboutness* is simply a different notion from *relevant*.

6 Conclusion

We have presented an investigation of the notion of relevance with respect to weak or subjunctive conditionals. The intent has been to characterise this notion with respect to the class of 'commonsense' or weak conditionals as expressed within the general framework of conditional logics. These logics have been used to represent counterfactuals, default properties, obligation, and other subjunctive conditions. Consequently, the account given here is applicable in a substantially broader class of settings than are previous accounts.

We have argued that, in this framework, relevance is not a relation between two arbitrary propositions, but rather, more narrowly, is a relation between a proposition and a weak conditional. In addition, it appears that any adequate definition of relevance must be stated at the metalevel. Our approach has been to propose an initial naïve definition and then use this definition to further examine relevance. A series of postulates characterising relevance was presented and, from this, an explicit definition for relevance was given.

As mentioned at the outset, we feel that 'relevant' is a concept for which we have no deep understanding. Our primary goal, therefore, has been to attempt to better understand the concept of relevance. However, this approach may have some practical consequences. Consider again the underlying framework: 'relevant' is a concept expressed in the metatheory of a conditional logic. Typically in such logics one does not have a principle of modus ponens for the weak conditional. For example, given that birds normally fly and that a particular individual is a bird, we cannot logically conclude that this individual flies. Yet the consequent of the conditional constitutes a plausible *default* conclusion. Hence, all other things being equal, a bird may (pragmatically) be concluded to fly.

What the present approach may allow then is a uniform means of sanctioning default inferences in a wide class of logics. Again, given that an individual is a green bird, we cannot conclude within the logic that this individual flies. However we may be able to argue in the metatheory that there are no provably relevant conditions 'blocking' the conclusion that the bird flies, and so draw this conclusion in the object language. Thus, in general we might realise a default reasoning system in the metatheory by allowing that, if β is contingently true and a conditional $\beta \Rightarrow \gamma$ is true, and there are no provably relevant conditions with respect to $\beta \Rightarrow \gamma$, then one can conclude γ by default. Hence default reasoning is defined in terms of relevance, or, in other words, we may *reduce* default reasoning to notions of relevance. The advantage of this programme, clearly, is that it would provide a uniform, justified approach to default reasoning in a wide class of logics, and consequently for a wide range of applications.

Acknowledgements

This work began at a workshop on relevance held at Kananaskis, Alberta; we thank the other participants for their helpful and instructive suggestions: Fahiem Bacchus, Craig Boutilier, Randy Goebel, Hector Levesque, Bob Mercer, Eric Neufeld, David Poole, and Ray Reiter. This research was supported in part by the Natural Science and Engineering Research Council of Canada and the Institute for Robotics and Intelligent Systems (IRIS) in the Networks of Centres of Excellence Program of Canada.

Appendix

A Proofs of Theorems

Lemma A.1 *The following formulae are valid.*

CSO: $(\alpha \Rightarrow \beta \wedge \beta \Rightarrow \alpha) \implies (\alpha \Rightarrow \gamma \equiv \beta \Rightarrow \gamma)$

RT: $\alpha \wedge \beta \Rightarrow \gamma \implies (\alpha \Rightarrow \beta \implies \alpha \Rightarrow \gamma)$

CMP: $\alpha \wedge \beta \Rightarrow \gamma \implies \alpha \Rightarrow (\beta \implies \gamma)$

ST: $((\beta \Rightarrow \gamma) \wedge (\alpha \wedge \beta \Rightarrow \neg\gamma)) \implies \beta \Rightarrow \neg\alpha$

Proof: Since the following results are elementary, we have on occasion combined several steps into one.

CSO: [Bur81]

RT:

1. $\alpha \wedge \beta \Rightarrow \alpha$ — [Bur81]/RI
2. $\alpha \Rightarrow \beta \implies \alpha \Rightarrow \alpha \wedge \beta$ — ID,CC,PC
3. $((\alpha \wedge \beta \Rightarrow \alpha) \wedge (\alpha \Rightarrow \alpha \wedge \beta)) \implies (\alpha \wedge \beta \Rightarrow \gamma \equiv \alpha \Rightarrow \gamma)$ — CSO
4. $\alpha \Rightarrow \beta \implies (\alpha \wedge \beta \Rightarrow \gamma \equiv \alpha \Rightarrow \gamma)$ — 1,2,3,PC
5. $\alpha \wedge \beta \Rightarrow \gamma \implies (\alpha \Rightarrow \beta \implies \alpha \Rightarrow \gamma)$ — 4,PC

CMP:

1. $\alpha \wedge \beta \Rightarrow \gamma$ — Supp.
2. $\alpha \wedge \beta \Rightarrow \neg\beta \vee \gamma$ — 1,Easy result
3. $\alpha \wedge \neg\beta \Rightarrow \neg\beta$ — [Bur81]/RI
4. $\alpha \wedge \neg\beta \Rightarrow \neg\beta \vee \gamma$ — 3,Easy result
5. $\alpha \Rightarrow \neg\beta \vee \gamma$ — 2,4,CA
6. $\alpha \wedge \beta \Rightarrow \gamma \implies \alpha \Rightarrow (\beta \implies \gamma)$ — 1,5,Disch.,Subst Equiv

ST:

1. $\beta \Rightarrow \gamma$ — Supp.
2. $\alpha \wedge \beta \Rightarrow \neg\gamma$ — Supp.
3. $\alpha \wedge \beta \Rightarrow \neg\gamma \implies \beta \Rightarrow (\alpha \implies \neg\gamma)$ — CMP
4. $\beta \Rightarrow (\alpha \implies \neg\gamma)$ — 2,3,MP
5. $\beta \Rightarrow (\gamma \wedge (\gamma \implies \neg\alpha))$ — 1, 4, CC

6. $\beta \Rightarrow (\gamma \wedge \neg\alpha)$ 5, Subst Equiv.
7. $\beta \Rightarrow \neg\alpha$ 6,CM
8. $((\beta \Rightarrow \gamma) \wedge (\alpha \wedge \beta \Rightarrow \neg\gamma)) \implies \beta \Rightarrow \neg\alpha$ 1,2,7,Disch.

■

Lemma A.2 *If α is relevant to $\beta \Rightarrow \gamma$ in $\mathcal{T}$ then*

1. $\mathcal{T} \models_L \beta \Rightarrow \gamma$
2. $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$ *or* $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \gamma$.

Proof: For the first part, if α is minimally relevant then the result is immediate. If α is relevant via part 2 of Definition 4.4, then we have that there is α' such that $\mathcal{T} \models \Box(\alpha \implies \alpha')$ where

1. α' is relevant to $\beta \Rightarrow \gamma$ and
2. α is relevant to $\alpha' \wedge \beta \Rightarrow \neg\gamma$.

Since $\mathcal{T}$ is finite, an inductive argument on 1. yields that there is a formula α'' that is minimally relevant to $\beta \Rightarrow \gamma$, from which the result follows immediately.

For part 2, an inductive argument based on Definition 4.4.2 yields that α is minimally relevant to a conditional of the form $\alpha' \wedge \alpha'' \wedge \cdots \wedge \beta \Rightarrow [\neg]\gamma$. Hence $\mathcal{T} \models_L \alpha \wedge \alpha' \wedge \alpha'' \wedge \cdots \wedge \beta \Rightarrow [\neg]\gamma$ or, since $\mathcal{T} \models_L \Box(\alpha \implies \alpha' \wedge \alpha'' \wedge \dots)$, we get $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow [\neg]\gamma$. ■

Theorem 4.1 Definition 4.4 satisfies **P0 – P4, P6, P8, PL0**.
Proof: P0 and **P1** are trivially satisfied.

For **P2**, assume that $\mathcal{T} \models_L \beta \Rightarrow \alpha$ and (contrary to what is to be shown) that α is relevant to $\beta \Rightarrow \gamma$. By Lemma A.2 we have that $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$ or $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \gamma$.

1. If $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$ then, since we have $\mathcal{T} \models_L \beta \Rightarrow \alpha$, we obtain $\mathcal{T} \models_L \beta \Rightarrow \alpha \wedge \beta$. Since we also have $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \beta$, using **CSO** we obtain that $\mathcal{T} \models_L \beta \Rightarrow \neg\gamma$. But from Lemma A.2 we have that $\mathcal{T} \models_L \beta \Rightarrow \gamma$, from which we obtain $\mathcal{T} \models \Box\neg\beta$, so $\mathcal{T} \models \Box\neg(\alpha \wedge \beta)$, contradicting Definition 4.3.
2. If $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \gamma$ then, since by assumption $\mathcal{T} \models_L \beta \Rightarrow \alpha$, we obtain by **RT** that $\mathcal{T} \models_L \beta \Rightarrow \gamma$. Hence by Definition 4.3, part 2, α is not minimally relevant to $\beta \Rightarrow \gamma$.

   So this means that (part 2. of Definition 4.4) there is a proposition α' where $\mathcal{T} \models \Box(\alpha \implies \alpha')$ and α' is relevant to $\beta \Rightarrow \gamma$. Moreover, since the set of relevant propositions is minimal, and $\mathcal{T}$ is finite, we can assume that $\mathcal{T} \not\models \Box(\alpha' \implies \alpha)$. The result then follows by an inductive argument based on the strength (wrt implication) of α – that is, that there is some α' that is minimally relevant to $\beta \Rightarrow \gamma$. But as shown in 1. above, this is impossible. (Note that since $\mathcal{T}$ is a finite propositional theory the induction is guaranteed to terminate.)

P3 is immediate since we disallow conditionals with antecedents that are necessarily false.

For **P4**, assume that $\mathcal{T} \models_L \Box(\alpha_1 \vee \cdots \vee \alpha_n)$ and for every α_k, $1 \leq k \leq n$, we have that α_k is relevant to $\beta \Rightarrow \gamma$. Thus $\mathcal{T} \models_L \beta \Rightarrow \gamma$ (Lemma A.2). But if every α_k is minimally relevant to $\beta \Rightarrow \gamma$ then we have that $\mathcal{T} \models_L \alpha_k \wedge \beta \Rightarrow \neg\gamma$ for every such α_k, whence by repeated **CA**, $\mathcal{T} \models_L (\bigvee_{i=1}^{n} \alpha_i) \wedge \beta \Rightarrow \neg\gamma$ or $\mathcal{T} \models_L \beta \Rightarrow \neg\gamma$. But $\mathcal{T} \models_L \beta \Rightarrow \gamma$, so $\mathcal{T} \models \Box\neg\beta$, and $\mathcal{T} \models \Box\neg(\alpha \wedge \beta)$, contradiction. On the other hand, if some α_k is relevant, but not minimally relevant, then there is (by an inductive argument on the strength of α_k) some α'_k such that α'_k is minimally relevant and $\mathcal{T} \models_L \Box(\alpha_k \implies \alpha'_k)$ but $\mathcal{T} \not\models_L \Box(\alpha'_k \implies \alpha_k)$, from which this case reduces to the preceding. (Again, since $\mathcal{T}$ is a finite propositional theory the induction is guaranteed to terminate.)

P6 follows immediately from part 2. of Definition 4.4.

For **P8** we prove the case where $n = 2$; the case for arbitrary n is a straightforward extension.

Assume that α, $\alpha \wedge \alpha_1$, and $\alpha \wedge \alpha_2$ are relevant to $\beta \Rightarrow \gamma$ where $\mathcal{T} \models_L \Box(\alpha_1 \vee \alpha_2)$.

Without loss of generality assume that $\mathcal{T} \models_L \beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$. (The cases for $\mathcal{T} \models_L \beta \Rightarrow \neg\gamma$ and $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \gamma$ follow analogously.)

There are four cases to consider:

1. $\mathcal{T} \models_L \alpha \wedge \alpha_1 \wedge \beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \alpha \wedge \alpha_2 \wedge \beta \Rightarrow \gamma$.
2. $\mathcal{T} \models_L \alpha \wedge \alpha_1 \wedge \beta \Rightarrow \neg\gamma$ and $\mathcal{T} \models_L \alpha \wedge \alpha_2 \wedge \beta \Rightarrow \gamma$.
3. $\mathcal{T} \models_L \alpha \wedge \alpha_1 \wedge \beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \alpha \wedge \alpha_2 \wedge \beta \Rightarrow \neg\gamma$.
4. $\mathcal{T} \models_L \alpha \wedge \alpha_1 \wedge \beta \Rightarrow \neg\gamma$ and $\mathcal{T} \models_L \alpha \wedge \alpha_2 \wedge \beta \Rightarrow \neg\gamma$.

1. Since $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$ and $\mathcal{T} \models_L \alpha \wedge \alpha_1 \wedge \beta \Rightarrow \gamma$ by **ST** we obtain $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\alpha_1$. Using the fact that $\mathcal{T} \models_L \Box(\alpha_1 \vee \alpha_2)$, we obtain that $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \alpha_2$ (i).

   But $\mathcal{T} \models_L \alpha \wedge \alpha_2 \wedge \beta \Rightarrow \gamma$ so by an analogous argument using **ST** we get $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\alpha_2$. But this together with (i) gives $\mathcal{T} \models_L \Box\neg(\alpha \wedge \beta)$, contradiction.

2. Since $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$ and $\mathcal{T} \models_L \alpha \wedge \alpha_1 \wedge \beta \Rightarrow \gamma$ by **ST** we obtain $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\alpha_1$, and so $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \alpha_2$.

   Since $\mathcal{T} \models_L (\alpha \wedge \beta \wedge \alpha_2) \Rightarrow (\alpha \wedge \beta)$, and since it follows by **ID** and **CC** that $\mathcal{T} \models_L (\alpha \wedge \beta) \Rightarrow (\alpha \wedge \beta \wedge \alpha_2)$, we obtain $\mathcal{T} \models_L (\alpha \wedge \beta) \Leftrightarrow (\alpha \wedge \beta \wedge \alpha_2)$.

   3. follows analogously.

4. Since α is relevant to $\beta \Rightarrow \gamma$, neither $\alpha \wedge \alpha_1$ nor $\alpha \wedge \alpha_2$ are minimally relevant, since this would contradict part 3 of Definition 4.3. So for $\alpha \wedge \alpha_1$, this means that there is α'_1 such that $\mathcal{T} \models \Box((\alpha \wedge \alpha_1) \implies \alpha'_1)$ where α'_1 is relevant to $\beta \Rightarrow \gamma$ and $\alpha \wedge \alpha_1$ is relevant to $\alpha'_1 \wedge \beta \Rightarrow \neg\gamma$. For $\alpha \wedge \alpha_2$, this means that there is α'_2 such that $\mathcal{T} \models \Box((\alpha \wedge \alpha_2) \implies \alpha'_2)$ where α'_2 is relevant to $\beta \Rightarrow \gamma$ and $\alpha \wedge \alpha_2$ is relevant to $\alpha'_2 \wedge \beta \Rightarrow \neg\gamma$. By the minimality of relevant conditions, we have $\mathcal{T} \not\models \Box(\alpha'_1 \implies (\alpha \wedge \alpha_1))$ and $\mathcal{T} \not\models \Box(\alpha'_2 \implies (\alpha \wedge \alpha_1))$. But now with the assertions:

α'_1 is relevant to $\beta \Rightarrow \gamma$ and
α'_2 is relevant to $\beta \Rightarrow \gamma$,

we have reduced the case at hand to case 1. above which we showed cannot occur.

For **PL0**, the contrapositive is easiest to show. If α is minimally relevant to $\beta \Rightarrow \gamma$ then the result is immediate. Otherwise, from part 2. of Definition 4.4, α is relevant to some strictly weaker conditional $\alpha' \wedge \beta \Rightarrow \neg\gamma$. The result follows by an inductive argument on the strength of the antecedent of the conditional, together with the Limit Assumption, which forbids indefinite successively-weaker antecedents. ■

Theorem 4.2 Definition 4.6 satisfies **P1 – P6, P8, PL0**.
Proof: P5 follows immediately from Definitions 4.5 and 4.6. The remaining principles follow as in Theorem 4.1, together with the observation that if α is whether-relevant to $\beta \Rightarrow \gamma$ then either $\mathcal{T} \models_L \beta \Rightarrow \gamma$ or $\mathcal{T} \models_L \beta \Rightarrow \neg\gamma$, but not both. ■

Lemma A.3 *Let $\mathcal{T}$ be a consistent theory such that $\mathcal{T} \models_L \beta \Rightarrow \gamma$ and $\mathcal{T} \models_L \alpha \wedge \beta \Rightarrow \neg\gamma$. Then there is an extension $\mathcal{T}'$ of $\mathcal{T}$ such that for some α' where $\mathcal{T} \models_L \Box(\alpha \implies \alpha')$ we have that α' is relevant to $\beta \Rightarrow \gamma$ in $\mathcal{T}'$.*

Proof: If α is not a basic proposition, then let $\mathcal{T}' = \mathcal{T} \cup \{\Box(p \equiv \alpha)\}$ where p is an atomic sentence not occurring in $\mathcal{T}$. So α is a basic proposition in $\mathcal{T}'$; so we can always extend $\mathcal{T}$ so that α is a basic proposition. If there is no α' where α' is a basic proposition, $\mathcal{T} \models_L \Box(\alpha \implies \alpha')$, $\mathcal{T} \not\models_L \Box(\alpha \equiv \alpha')$, and $\mathcal{T} \models_L \alpha' \wedge \beta \Rightarrow \neg\gamma$, then by Definition 4.3, α is minimally relevant to $\beta \Rightarrow \gamma$ in $\mathcal{T}'$. If there is such an α', then either α' is minimally relevant to $\beta \Rightarrow \gamma$ in $\mathcal{T}'$, or else the process can be repeated to find some α' satisfying the above conditions. Since $\mathcal{T}$ is finite, this process is guaranteed to terminate, and so we will find some condition minimally relevant to $\beta \Rightarrow \gamma$ in $\mathcal{T}'$. ■

Theorem 4.3 Relevance is non-trivial; that is, there are theories with extant relevance conditions.
Proof: In any interesting conditional logic there will be consistent theories in which sentences of the form $\beta \Rightarrow \gamma$ and $\alpha \wedge \beta \Rightarrow \neg\gamma$ are true. According to Lemma A.3 either there will be a trivial extension of the theory in which α is relevant, or for some α' where $\mathcal{T} \models_L \Box(\alpha \implies \alpha')$, we have that α is relevant to $\beta \Rightarrow \gamma$ in the theory. ■

References

[AB75] A.R. Anderson and N.D. Belnap Jr. *Entailment: The Logic of Relevance and Necessity, Vol. I.* Princeton University Press, 1975.

[Bou94] C. Boutilier. Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68(1):87–154, 1994.

[Bur81] J.P. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22(1):76–84, 1981.

[Del87] J.P. Delgrande. A first-order conditional logic for prototypical properties. *Artificial Intelligence*, 33(1):105–130, 1987.

[DFHP94] D. Dubois, L. Fariñas del Cerro, A. Herzig, and H. Prade. An ordinal view of independence with application to plausible reasoning. In *Uncertainty in AI'94*, Seattle, WA, July 1994.

[DP94] A. Darwiche and J. Pearl. Symbolic causal networks. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 238–244, Seattle, WA, 1994.

[Gär78] P. Gärdenfors. On the logic of relevance. *Synthese*, 37(3):351–367, 1978.

[Gär88] P. Gärdenfors. *Knowledge in Flux: Modelling the Dynamics of Epistemic States.* The MIT Press, Cambridge, MA, 1988.

[Goo61] N. Goodman. About. *Mind*, 70:1–24, 1961.

[Goo73] N. Goodman. *Fact, Fiction, and Forecast.* Bobbs Merrill Co., Inc., 1973.

[HC68] G.E. Hughes and M.J. Cresswell. *An Introduction to Modal Logic.* Methuen and Co. Ltd., 1968.

[Lam91] Philippe Lamarre. S4 as the conditional logic of nonmonotonicity. In *Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning*, pages 357–367, Cambridge, MA, April 1991.

[Lew73] D. Lewis. *Counterfactuals.* Harvard University Press, 1973.

[Lov78] David W. Loveland. *Automated Theorem Proving: A Logical Basis.* North-Holland Inc., N.Y., 1978.

[LS93] A. Y. Levy and Y. Sagiv. Exploiting irrelevance reasoning to guide problem solving. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 138–144, Chambéry, Fr., 1993.

[McC86] J. McCarthy. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[Nil80] N.J. Nilsson. *Principles of Artificial Intelligence.* Tioga Pub. Co., 1980.

[Nut80] D. Nute. *Topics in Conditional Logic*, volume 20 of *Philosophical Studies Series in Philosophy*. D. Reidel Pub. Co., 1980.

[Pea88] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Mateo, CA, 1988.

[Pea90] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In R. Parikh, editor, *Proc. of the Third Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 121–135, Pacific Grove, Ca., 1990. Morgan Kaufmann Publishers.

[Sch86] G. N. Schlesinger. Relevance. *Theoria*, 52:57–67, 1986.

[SG87] D. Subramanian and M. R. Genesereth. The relevance of irrelevance. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 416–422, Milan, Italy, 1987.

[Sta68] R.C. Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in Logical Theory*, pages 98–112. Basil Blackwell, Oxford, 1968.