

# Supraclassical Inference without Probability

*David Makinson*

Chapter 6 (pp 95-111) in P. Bourguine & J-P. Nadal eds, *Cognitive Economics: An Interdisciplinary Approach*. Springer Verlag, 2003.

## 1. Introduction

### 1.1. Abstract

A logic is said to be supraclassical if it permits us to infer more from a set of premises than classical logic authorizes. It is called nonmonotonic if increasing the premises available for inference can lead to loss, as well as gain, of conclusions.

Probabilistic inference is a well-known kind of nonmonotonic supraclassical reasoning, and is familiar to many economists. But there are qualitative kinds as well, little known outside circles of logicians and computer scientists. They allow us to conclude more than classical logic permits, without appeal to probability distributions. Like probabilistic inference, they are also nonmonotonic.

The purpose of this paper is to take some of the mystery out of these systems, by showing that they are not as unfamiliar as may at first sight appear. In fact, they are easily accessible to anybody with a background in classical propositional logic.

### 1.2. Recalling classical consequence

We refresh the reader's memory with the basic notions of classical propositional consequence, which are essential for what is to follow.

Classical logic uses a formal language whose propositions (or formulae – we will use the two terms interchangeably) are made up from an infinite list of elementary letters  $q, r, s, \dots$  by means of the two-place connectives  $\wedge, \vee$  and the one-place connective  $\neg$ , understood in terms of their usual truth-tables, respectively for conjunction, disjunction, and negation. Other truth-functional connectives such as material implication  $\rightarrow$  are defined from these three in the usual ways:  $a \rightarrow x$  abbreviates  $\neg(a \wedge \neg x)$  and  $a \leftrightarrow x$  abbreviates  $(a \rightarrow x) \wedge (x \rightarrow a)$ . Such formulae are called *Boolean*. Note that although infinitely many elementary letters are available, each formula may contain only a finite number of them. Likewise, we can apply any of the three connectives any finite number of times, but we do not allow infinite conjunctions or disjunctions. Each formula is thus finitely long.

An *assignment* is a function on the set of all elementary letters into the two-element set  $\{1,0\}$ . Each assignment may be extended in a unique way to a *valuation*, that is, a function  $v$  on the set of all formulae into the two-element set  $\{1,0\}$  that agrees with the assignment on elementary letters and behaves in accord with the standard truth-tables for the compound formulae made up using  $\wedge, \vee, \neg$ . That is,  $v(a \wedge b) = 1$  iff  $v(a) = v(b) = 1$ ,  $v(a \vee b) = 0$  iff  $v(a) = v(b) = 0$ , and  $v(\neg a) = 1$  iff  $v(a) = 0$ . When  $A$  is a *set* of formulae, one writes  $v(A) = 1$  as shorthand for  $v(a) = 1$  for all  $a \in A$ .

Let  $A$  be any set of formulae, and let  $x$  be an individual formula. One says that  $x$  is a *classical consequence* of  $A$  iff there is no valuation  $v$  such that  $v(A) = 1$  whilst  $v(x) = 0$ . The standard notation is  $A \vdash x$ , and the sign  $\vdash$  is called ‘gate’ or ‘turnstile’. When dealing with individual formulae on the left, the notation is simplified a little by dropping parentheses, writing  $a \vdash x$  in place of  $\{a\} \vdash x$ .

Thus classical consequence is a *relation* between propositions, or more generally between *sets*  $A$  of such propositions on the left and individual propositions  $x$  on the right. It may also be seen as an *operation* acting on sets  $A$  of propositions to give larger sets  $Cn(A)$ . These two representations of classical consequence are trivially interchangeable. Given a relation  $\vdash$ , we may define the operation  $Cn$  by setting  $Cn(A) = \{x: A \vdash x\}$ ; and conversely we may define  $\vdash$  from  $Cn$  by the rule  $A \vdash x$  iff  $x \in Cn(A)$ .

Both of the representations are useful. Sometimes one is more convenient than another. For example, it is often easier to visualize things in terms of the relation, but more concise to formulate and prove them using the operation. The same will be true when we come to non-classical consequence. For this reason, in this paper we will constantly be hopping from one notation to the other, as two ways of saying the same thing, and we encourage the reader to do the same.

Classical consequence has a number of very useful properties. To begin with, it is a *closure relation*, in the sense that it satisfies the following three conditions for all formulae  $a, x$  and all sets  $A, B$  of formulae:

<i>Reflexivity</i> alias <i>Inclusion</i>	$A \vdash a$ whenever $a \in A$
Cumulative Transitivity, CT (alias <i>Cut</i> )	Whenever $A \vdash b$ for all $b \in B$ and $A \cup B \vdash x$ then $A \vdash x$
<i>Monotony</i>	Whenever $A \vdash x$ and $A \subseteq B$ then $B \vdash x$

Expressed in the language of  $Cn$ , this means that classical consequence is a *closure operation* in the sense that it satisfies the following conditions.

<i>Reflexivity</i>	$A \subseteq Cn(A)$
<i>Cumulative Transitivity</i>	$A \subseteq B \subseteq Cn(A)$ implies $Cn(B) \subseteq Cn(A)$
<i>Monotony</i>	$A \subseteq B$ implies $Cn(A) \subseteq Cn(B)$

The three conditions defining the notion of a closure relation are examples of what are known as *Horn rules*. Roughly speaking, a Horn rule tells us that if such-and-such and so-and-so (any number of times) are all elements of the relation, then so is something else. None of the suppositions of a Horn rule may be negative – none of them can require that something is *not* an element of the relation. Nor is the conclusion allowed to be disjunctive – it cannot say that given the suppositions, *either* this *or* that is in the relation. Horn rules have very useful properties, most notably that whenever a Horn rule is satisfied by every relation in a family, then the relation formed by taking the intersection of the entire family also satisfies the rule.

Finally, we recall that classical consequence is *compact*, in the sense that whenever  $A \vdash x$  then there is a finite subset  $A' \subseteq A$  with  $A' \vdash x$ . In the language of operations, whenever  $x \in Cn(A)$  then there is a finite subset  $A' \subseteq A$  with  $x \in Cn(A')$ .

These are *abstract* properties of classical consequence, in the sense that they make no reference to any of the connectives  $\wedge, \vee, \neg$ . Evidently, the relation also has a number of properties concerning each of these connectives arising from their respective truth-tables, for example the property that  $a \vee b \in Cn(a)$ . We shall not enumerate these, but recall just one that will play a very important role in what follows: the property of *disjunction in the premises*, *alias* OR. It says that whenever  $A \cup \{a\} \vdash x$  and  $A \cup \{b\} \vdash x$  then  $A \cup \{a \vee b\} \vdash x$ . In the language of operations,  $Cn(A \cup \{a\}) \cap Cn(A \cup \{b\}) \subseteq Cn(A \cup \{a \vee b\})$ .

### 1.3. Using probability to get more conclusions out of your premises

Is there any way of obtaining more conclusions from a given set of premises than is authorized by the canons of classical logic? In particular, is there any way of doing this within the confines of a simple Boolean language like that for classical propositional logic described above?

To be sure, any additional conclusion will not be an *ineluctable* consequence of the premises. For since it is not a classical consequence in the sense that we have defined, there will be some combination of truth-values for the elementary letters occurring within it, that makes the premises all true but the conclusion false. Nevertheless, there may be conclusions that we feel can *reasonably* be derived. But by what kind of procedure?

Probability theory already provides us with one such procedure. In particular, the notion of *threshold probability consequence* permits us to obtain all classically valid conclusions and more.

We recall the definition. Let  $P$  be any set of probability distributions  $p$  on the Boolean language  $L$  into the real interval  $[0,1]$ . Let  $t \in [0,1]$ . We call  $t$  the *threshold*. Probabilistic inference (more fully, threshold probabilistic inference) is

defined as follows: for any propositions  $a, x$  we put  $a \mid\sim_{P,t} x$  iff for all  $p \in P$ , if  $p(a) \neq 0$  then  $p(a \wedge x)/p(a) \geq t$ .

Thus there is not one probabilistic inference relation  $\mid\sim_{P,t}$  but a family of them, one for each set  $P$  of probability distributions and each choice of a threshold value  $t$ . There is some discussion concerning the most appropriate ways of extending the definition so as to cover infinite premise sets as well as single-proposition premises, but we will leave that aside.

The first point to observe is that threshold probabilistic inference relations are *supraclassical*. That is, it is easy to show that whenever  $a \mid\sim_{P,t} x$  for Boolean formulae  $a, x$  then for any set  $P$  of probability distributions and any threshold  $t$ , we have  $a \mid\sim_{P,t} x$ . In other words, whenever  $a \mid\sim_{P,t} x$  fails, there is an assignment  $v$  of truth-values to the elementary letters occurring in  $a, x$  such that  $v(a) = 1$  and  $v(x) = 0$ .

The second point is that they are *nonmonotonic*. We may have  $a \mid\sim_{P,t} x$  but not  $a \wedge b \mid\sim_{P,t} x$ . It is easy to choose appropriate Boolean formulae  $a, b, x$  and define a suitable probability distribution  $p$  such that  $p(a \wedge x)/p(a) \geq t$  but  $p(a \wedge b \wedge x)/p(a \wedge b) < t$ , so that choosing  $P = \{p\}$ , we have  $a \mid\sim_{P,t} x$  but not  $a \wedge b \mid\sim_{P,t} x$ .

Suppose for example that the language has just two elementary letters  $q, r$ . Consider the probability distribution that gives each of the four atoms  $q \wedge r, \dots, \neg q \wedge \neg r$  equal values 0.25, and choose the threshold value  $t = 0.5$ . Put  $a = q \vee \neg q, x = q \vee r, b = \neg q \wedge \neg r$ . Then  $p(a \wedge x)/p(a) = 0.75 \geq t$  while  $p(a \wedge b \wedge x)/p(a \wedge b) = 0 < t$ .

While the relation of probabilistic consequence is undoubtedly useful, logicians have also sought for other ways of going beyond the limits of classical consequence within the basic Boolean language. They have done this for at least two reasons.

- An *epistemic* reason: There are many situations in ordinary life in which we have no rational basis for assuming a specific probability distribution or restricted set of them, but where we still feel that it is possible to infer with reasonable confidence beyond the limits of classical consequence.
- A *pragmatic* reason: Even when we do have grounds for assuming a specific probability distribution, or restricted set of them, the calculations involved may be such that it is not feasible to calculate answers to questions “does this imply that?” (for a given threshold  $t$ ) within the time in which the answer could be useful.

There is another kind of reason or a more *theoretical*, even *aesthetic*, nature. From the point of view of the logician, threshold probabilistic inference is very badly behaved.

- Threshold probabilistic inference fails not only monotony, but also certain other basic Horn conditions that are usually regarded as desirable.

In particular, it fails the principle of *conjunction of conclusions*. That is, we may have  $a \sim_{p,t} x$  and  $a \sim_{p,t} y$  but not  $a \sim_{p,t} x \wedge y$ . In the same example as above, put  $y = \neg q$ . Then  $a \sim_{p,t} x$  since again  $p(a \wedge x)/p(a) = 0.75 \geq t$ , and  $a \sim_{p,t} y$  since  $p(a \wedge y)/p(a) = 0.5 \geq t$ , but not  $a \sim_{p,t} x \wedge y$  since  $p(a \wedge x \wedge y)/p(a) = 0.25 < t$ .

Essentially because of the failure of conjunction of conclusions, probabilistic consequence operations also fail the principle of *cumulative transitivity* alias *cut* mentioned in the preceding section. Indeed, this can happen for singleton sets of formulae and a singleton set of probability distributions. In other words, it can happen that  $a \sim_{p,t} x$  and  $a \wedge x \sim_{p,t} y$  but not  $a \sim_{p,t} y$ . Plain transitivity also fails, for we can have  $a \sim_{p,t} x$  and  $x \sim_{p,t} y$  without  $a \sim_{p,t} y$ .

For all these reasons, then, logicians have developed a number of systems that enable us to conclude more than classical logic allows, but without any use of probabilities or, indeed, of any kind of function taking formulae into the real interval. The systems are thus *qualitative* in nature. In the jargon, they are known as systems of *nonmonotonic logic*. The choice of name comes from the fact that the inference relations developed are indeed nonmonotonic. It is however rather misleading because as we have just seen, probabilistic inference is also nonmonotonic. But that is the standard name, and there is no point in trying to change it in this paper.

The literature on nonmonotonic logics is rather difficult of access. There are a great many different such systems, and the reader can easily lose bearings trying to find a way among them. The language of the logicians is also often unfamiliar to economists.

The purpose of this paper is to take some of the mystery out of these systems, by showing that they are not as unfamiliar as may at first sight appear. In fact, they are easily accessible to anybody with a background in classical propositional logic – a full understanding of section 1.2 above should in principle suffice. If the reader falls into difficulty at any stage, he should go back to that section to resolve it.

In order to render the nonmonotonic systems as transparent as possible, we will show how there are natural ‘bridges’ between classical consequence and the principal kinds of nonmonotonic logic to be found in the literature. Like classical logic, they are perfectly monotonic, but they already display some of the distinctive features of the nonmonotonic systems. As well as providing easy conceptual passage to the nonmonotonic case these logics have an interest of their own.

#### 1.4. Three qualitative ways of getting more conclusions

We will describe three different ways, each of them qualitative, of getting out of a set of premises more than is authorized by straightforward application of classical consequence. Roughly speaking, the first method uses additional

*background assumptions*. The second restricts the set of *valuations* that are considered possible. And the third uses additional *background rules*.

Each of these procedures gives rise to a corresponding kind of monotonic consequence relation. They are not entirely equivalent to each other. But they all give us supraclassical closure operations, i.e. operations that include classical consequence and satisfy reflexivity, cut and monotony. We call such consequence relations *paraclassical*.

The three kinds of paraclassical consequence serve as conceptual paths to corresponding families of nonmonotonic consequence, formed essentially by allowing key elements of the respective constructions to vary with the premises under consideration.

The situation is represented in Figure 1. The glorious sun of classical consequence illuminates the firmament from its centre. Three kinds of paraclassical consequence circle around it like planets: *pivotal-assumption*, *pivotal-valuation*, and *pivotal-rule* consequence. Their key ingredients are, respectively, a set of additional background assumptions, a reduced set of valuations, and an additional set of rules. By allowing these ingredients to vary in a principled manner with the premises of any given inference, we obtain three satellite kinds of nonmonotonic consequence operation: *default-assumption*, *default-valuation*, and *default-rule* consequence.

## 2. First Path - Using Additional Background Assumptions

### 2.1. From Classical Consequence to Pivotal Assumptions

We begin by examining the simplest kind of paraclassical consequence and its transformation into a form of nonmonotonic reasoning, namely inference with additional background assumptions.

In daily life, the assumptions that we make when reasoning are not all of the same level. Generally, there will be a few that we display explicitly, because they are special to the situation under consideration or in some other way deserving particular attention. There will usually be many others that we do not bother even to mention, because we take them to be part of shared common knowledge, or in some other way trivial. This phenomenon was already known to the ancient Greeks, who used the term *enthymeme* to refer to an argument in which one or more premises are left implicit. That is the idea that we develop in this section.

We work with the same propositional language as in classical logic, with the set of all its formulae called  $L$ . Let  $K \subseteq L$  be a set of formulae. Intuitively  $K$  will be playing the role of a set of background assumptions. Let  $A$  be any set of formulae, and let  $x$  be an individual formula.

We say that  $x$  is a *consequence of  $A$  modulo the assumption set  $K$* , and write  $A \vdash_K x$  *alias*  $x \in Cn_K(A)$  iff there is no valuation  $v$  such that  $v(K \cup A) = 1$  whilst  $v(x) = 0$ . Equivalently, iff  $K \cup A \vdash x$ . As simple as that! And we call a relation or operation a *pivotal-assumption consequence* iff it is identical with  $\vdash_K$  (resp.  $Cn_K$ ) for some set  $K$  of formulae. Note that there is not a unique pivotal-assumption consequence relation, but many – one for each value of  $K$ .

Since classical consequence is monotonic, pivotal-assumption consequence relations and operations are supraclassical in the sense defined earlier. That is, for every  $K$  we have  $\vdash \subseteq \vdash_K$  and  $Cn \leq Cn_K$ . They also share a number of abstract properties with classical consequence. In particular, as is immediate from the definition, they satisfy inclusion, cumulative transitivity and monotony, and thus are closure operations. They are also compact, and have the property of disjunction in the premises.

A striking feature of pivotal-assumption consequence (which separates it from the next bridge system that we will be describing) is that the above properties also suffice to characterize it. In other words, we have the following representation theorem for pivotal-assumption consequence. *Let  $\vdash'$  be any supraclassical closure relation that is compact and satisfies the condition of disjunction in the premises. Then there is a set  $K$  of formulae such that  $\vdash' = \vdash_K$ .*

## 2.2. From Pivotal Assumptions to Default Assumptions

In the definition of pivotal-valuation consequence, the set  $K$  of background assumptions was fixed through all variations of the premises  $A$  and potential conclusion  $x$ . If we allow  $K$  to vary with  $A$ , we obtain a different kind of inference relation. In particular, if we diminish  $K$  just as much as is necessary in order to preserve consistency with the current premises, we pass into the realm of nonmonotonicity.

More specifically, when the current premise set  $A$  is inconsistent with  $K$  we do not work with the whole of  $K$  but rather with the maximal subsets  $K'$  of  $K$  that are consistent with  $A$ . In general, there will be many such  $K'$ . We accept as output only what is common to all their separate outputs when taken together with  $A$ . We call this relation *default-assumption consequence*.

To give the definition explicitly, let  $K \subseteq L$  be a set of formulae, which again plays the role of a set of background assumptions. Let  $A$  be any set of formulae, and let  $x$  be an individual formula.

- We say that a subset  $K'$  of  $K$  is consistent with  $A$  iff there is a classical valuation  $v$  with  $v(K' \cup A) = 1$ .
- A subset  $K'$  of  $K$  is called *maximally consistent* (more briefly, *maxconsistent*) with  $A$  iff it is consistent with  $A$  but is not a proper subset of any  $K'' \subseteq K$  that is consistent with  $A$ .

- Finally, we define the relation  $\sim_K$  of *consequence modulo the default assumptions*  $K$  by putting  $A \sim_K x$  iff  $K' \cup A \vdash x$  for every subset  $K' \subseteq K$  that is maxiconsistent with  $A$ . Writing  $C_K$  for the corresponding operation, this puts  $C_K(A) = \bigcap \{Cn(K' \cup A) : K' \subseteq K \text{ and } K' \text{ maxiconsistent with } A\}$ .

We call a relation or operation a *default-assumption consequence* iff it is identical with  $\sim_K$  (resp.  $C_K$ ) for some set  $K$  of formulae. Note again that there is not a unique default-assumption consequence relation, but many – one for each value of  $K$ .

Default-assumption consequence operations/relations are nonmonotonic. That is, we may have  $A \sim_K x$  but not  $A \cup B \sim_K x$  where  $A, B$  are sets of propositions. Likewise, we may have  $a \sim_K x$  without  $a \wedge b \sim_K x$  where  $a, b$  are individual propositions.

To illustrate the failure of monotony, suppose  $K = \{q \rightarrow r, r \rightarrow s\}$  where  $q, r, s$  are distinct elementary letters of the language and  $\rightarrow$  is the truth-functional (*alias* material) conditional connective. Then  $q \sim_K s$  since the premise  $q$  is consistent with the whole of  $K$  and clearly  $\{q\} \cup K \vdash s$ . But  $q \wedge \neg r \not\sim_K s$ , since the premise  $q \wedge \neg r$  is no longer consistent with the whole of  $K$ . There is a unique maximal subset  $K' \subseteq K$  that is consistent with  $q \wedge \neg r$ , and that is  $K' = \{r \rightarrow s\}$ , and clearly  $\{q \wedge \neg r\} \cup K' \not\vdash s$  (witness the valuation  $v$  with  $v(q) = 1$  and  $v(r) = v(s) = 0$ ). We gained premises, but lost assumptions because of the consistency requirement.

As monotony can fail, default-assumption consequence operations are not in general closure operations. However they do satisfy both inclusion and cumulative transitivity (and hence also idempotence). They are also supraclassical and satisfy disjunction in the premises.

We note one further property. Although they may fail monotony, they always satisfy a weakened version of it called *cautious monotony*. This says that whenever  $A \sim_K x$  and  $A \sim_K y$  then  $A \cup \{x\} \sim_K y$ . More generally: whenever  $A \sim_K x$  for all  $x \in B$  and  $A \sim_K y$  then  $A \cup B \sim_K y$ . In the succinct notation of operations: whenever  $A \subseteq B \subseteq C_K(A)$  then  $C_K(A) \subseteq C_K(B)$ . We omit the verification; it can be found for example in Makinson (1994), which reviews the behaviour of these operations in considerable detail.

In summary, monotonic pivotal-assumption consequences  $Cn_K$  provide a natural ‘half-way house’ between classical  $Cn$  and nonmonotonic default-assumption consequences  $C_K$ . The definition of  $Cn_K$  adds the component  $K$  of background assumptions. The default operation requires as well the consistency of these assumptions with the set  $A$  of explicit premises, failing which it contracts from the assumptions to ensure consistency.

### 3. Second Path - Restricting the Set of Valuations

#### 3.1. From Classical Consequence to Pivotal Valuations

In the preceding section, we built a mid-station between classical and nonmonotonic reasoning. The essential idea was to augment the explicit premises  $A$  by a set  $K$  of background assumptions. We now do almost the same thing in a different way, which points towards a substantially different nonmonotonic generalization. We *restrict the set of valuations* that are considered. In other words, we fix a pivotal set  $W$  of valuations, and redefine consequence modulo it instead of the set  $V$  of all valuations.

Let  $W \subseteq V$  be a set of valuations on the language  $L$ . Let  $A$  be any set of formulae, and let  $x$  be an individual formula. We say that  $x$  is a *consequence of  $A$  modulo the valuation set  $W$* , and write  $A \vdash_W x$  *alias*  $x \in Cn_W(A)$  iff there is no valuation  $v \in W$  such that  $v(A) = 1$  whilst  $v(x) = 0$ . We call a relation or operation a *pivotal-valuation consequence* iff it coincides with  $\vdash_W$  (resp.  $Cn_W$ ) for some set  $W$  of valuations. Note again that there is not a unique pivotal-valuation consequence relation, but many – one for each value of  $W$ .

Immediately from this definition, pivotal-valuation consequence relations/operations are also supraclassical, i.e.  $Cn \leq Cn_W$  for any choice of  $W$ . They also satisfy inclusion, cumulative transitivity and monotony, and thus are closure operations. We are thus still in the realm of paraclassical inference. Pivotal-valuation consequence relations also have the property of disjunction in the premises. But, as we would already expect, they are not closed under substitution. Moreover - and this is the new feature - they are not always compact.

To illustrate the failure of compactness, let  $W$  be the set of all those valuations  $v$  such that only finitely many elementary letters (out of our infinite supply  $Q$  of elementary letters for generating formulae of the language  $L$ ) are true under  $v$ . On the one hand, for no finite subset  $Q' \subseteq Q$  do we have  $Q' \vdash_W (q \wedge \neg q)$ , since  $W$  contains a valuation that makes all the finitely many letters in  $Q'$  true, and this valuation evidently makes  $(q \wedge \neg q)$  false. On the other hand,  $Q \vdash_W (q \wedge \neg q)$  holds vacuously, since there is no valuation in  $W$  that satisfies all the infinitely many letters in  $Q$ .

This failure of compactness for some pivotal-valuation consequence operations shows that not every such operation is a pivotal-assumption one, for as we have seen, all of the latter do satisfy compactness. On the other hand, we do have the converse: every pivotal-assumption consequence operation is a pivotal-valuation one, so that the former family is strictly narrower than the latter family. The verification is almost immediate. Let  $Cn_K$  be any pivotal-assumption consequence operation with assumption set  $K$ . We need to show that  $Cn_K = Cn_W$  for an appropriate choice of a set  $W$  of valuations. It suffices to find a  $W \subseteq V$  such that for all  $A$  and all  $v \in V$ ,  $v(K \cup A) = 1$  iff  $v \in W$  and  $v(A) = 1$ . We simply put  $W$  be

the set of all valuations  $v$  such that  $v(K) = 1$ . Then for every valuation  $v$ ,  $v(K \cup A) = 1$  iff  $v(K) = 1$  and  $v(A) = 1$ , i.e. iff  $v \in W$  and  $v(A) = 1$  and we are done.

Putting this together with points already noted, we can conclude that *the pivotal-assumption consequence operations are precisely the pivotal-valuation ones that are compact*. For we have just shown that every pivotal-assumption consequence operation is a pivotal-valuation one, and we have already remarked that the former are always compact. Conversely, we have noted that any pivotal-valuation operation is a supraclassical consequence operation satisfying disjunction in the premises; so if it is also compact the representation theorem of section 2.1 tells us that it is a pivotal-assumption relation.

In particular, in a finite language (that is, one generated by boolean connectives from a finite set of elementary letters), where compactness always holds, the families of pivotal-assumption and pivotal-valuation consequence operations coincide. For the computer scientist, who always works in a finite language, pivotal-assumption and pivotal-valuation are thus equivalent. For the logician, who takes perverse pleasure in the subtleties of the infinite, they are not.

The question arises whether we can characterize the family of pivotal-valuation consequence operations in terms of the properties that these operations satisfy. It would be pleasant to be able to report that the family is fully characterized by the properties of being a closure operation, supraclassical, and satisfying disjunction in the premises. This would give us a representation theorem for the family, paralleling the one that we presented in section 2.1 for pivotal-assumption consequence operations. The theorem holds when the language is finite, since it holds for pivotal-assumption consequence. However in the infinite case Karl Schlechta has found a counterexample: a supraclassical closure operation satisfying disjunction in the premises that is not a pivotal-valuation consequence operation.

The question thus remains: can we characterize the family of all pivotal-valuation consequence operations in some other way, i.e. as the supraclassical closure operations satisfying disjunction in the premises plus some other property yet to be identified? As far as the author knows, this question has not been answered.

### 3.2. From Pivotal Valuations to Default Valuations

As the reader will already have guessed from the earlier discussion of pivotal-assumption consequence, nonmonotonicity can arise from pivotal valuations if we allow the restricted set  $W$  of valuations to vary with the premise set  $A$ .

The best-known way of doing this was devised by Shoham (1988). His essential idea was to focus on those valuations satisfying the premise set  $A$ , that are *minimal* under some background ordering  $<$  over valuations.

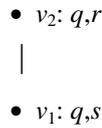
A *preferential model* is understood to be a pair  $(W, <)$  where  $W$  is, as in the monotonic case, a subset of the set  $V$  of all valuations on the language  $L$  and  $<$  is a relation over  $W$ . More generally, it is convenient to take  $W$  as an *indexed subset* (alias *multi-subset*) of  $V$ , thus allowing multiple copies of a single valuation, but that is a technical detail which we may leave aside.

Given a preferential model  $(W, <)$  we say that a formula  $x$  is a *preferential consequence* of a set  $A$  of formulae, and write  $A \sim_{<} x$  iff  $v(x) = 1$  for every valuation  $v \in W$  that is minimal among those in  $W$  that satisfy  $A$ . Thus the consequence relation depends on both  $W$  and  $<$ , and strictly speaking the snake sign should also carry  $W$  as a subscript; but we omit it to simplify notation. When preferential consequence is read as an operation rather than a relation, it is written as  $C_{<}(A)$ . Pivotal-valuation consequence  $A \vdash_W x$  thus becomes the special case where  $<$  is the empty relation so that all valuations in  $W$  are equally preferred.

It can be convenient to express the definition in a more compact manner. When  $W$  is a set of valuations and  $A$  is a set of formulae, write  $|A|_W$  for the set of all valuations in  $W$  that satisfy  $A$ , i.e.  $|A|_W = \{v \in W: v(A) = 1\}$ . Write  $\min_{<}(|A|_W)$  for the set of all minimal elements of  $|A|_W$ . In this notation,  $A \sim_{<} x$  iff whenever  $v \in \min_{<}(|A|_W)$  then  $v(x) = 1$ , i.e. iff  $\min_{<}(|A|_W) \subseteq |x|_W$ .

Note that in all these definitions, no constraints are placed on the relation  $<$  over  $W$ . It need not be transitive or even irreflexive – much less complete. The concept of minimality makes perfect sense for an arbitrary relation. If  $R$  is any relation whatsoever over a set  $X$ , then for any  $Y \subseteq X$  we take the minimal elements of  $Y$  under  $R$  to be the  $y \in Y$  such that for no  $y' \in Y$  do we have  $(y', y) \in R$ . The notion of a preferential model is thus well-defined without placing any constraints on its relation, but to guide intuitions it is useful to keep at the back of one's mind the typical case that it is both irreflexive and transitive.

Preferential consequence relations/operations are nonmonotonic. For a simple example, consider a language with at least three elementary letters  $q, r, s$ , and put  $W = \{v_1, v_2\}$  where  $v_1(q) = v_2(q) = 1$ ,  $v_1(r) = 0$ ,  $v_2(r) = 1$ ,  $v_1(s) = 1$ ,  $v_2(s) = 0$ , and order  $W$  by putting  $v_1 < v_2$ . Informally, we describe this by saying: let  $q$  be true in both valuations,  $r$  true in just the top one, and  $s$  true in just the bottom one. Diagrammatically:



Here we mention at each valuation point only the letters that are true there. Then  $q \sim_{<} s$  since the least valuation in which  $q$  is true is  $v_1$ , and  $s$  is also true there; but  $q \wedge r \not\sim_{<} s$ , since the least valuation in which  $q \wedge r$  is true is  $v_2$ , and  $s$  is false there.

Apart from failing monotony, preferential consequence relations are remarkably well behaved. They are supraclassical and satisfy disjunction in the premises. They also satisfy cumulative transitivity. However, they lack some of the properties of their assumption-based analogues. We mention two important ones.

Preferential consequence relations do not always satisfy cautious monotony, which is in effect, a converse of cumulative transitivity. In the language of consequence operations, cautious monotony says:  $A \subseteq B \subseteq C(A)$  implies  $C(A) \subseteq C(B)$ . In terms of relations: whenever  $A \sim b$  for all  $b \in B$  and  $A \sim x$  then  $A \cup B \sim x$ . In the case that  $A, B$  are singletons it amounts to: whenever  $a \sim b$  and  $a \sim x$  then  $a \wedge b \sim x$ .

Although this property may fail for preferential consequence relations, it always holds in the finite case. More generally, it holds whenever there are no infinite descending chains in the preferential model. More generally still, it holds whenever the model satisfies a condition known as *stopping* (alias *smoothness*). This says that whenever  $v \in |A|_W$  then either  $v \in \min_{<}(|A|_W)$  or there is a  $u < v$  with  $u \in \min_{<}(|A|_W)$ .

The main point that we want to emphasize in this section is that once again, a well-known kind of nonmonotonic consequence relation (preferential consequence) emerges from a monotonic one (pivotal-valuation consequence) by allowing its key ingredient (the set  $W$  of valuations allowed) to vary with the premise sets  $A$ , the variation being controlled by minimization under a relation. The monotonic operation thus serves as an antechamber to its nonmonotonic counterpart.

## 4. Third Path - Using Additional Rules

### 4.1. From Classical Consequence to Pivotal Rules

We come now to a third kind of system. The basic idea is similar to that of adding background assumptions, but instead of adding *propositions* we add *rules*. This apparently small difference brings with it considerable divergence, even in the finite case where (as noted towards the end of section 3.1) the pivotal-assumption and pivotal-valuation approaches are equivalent. Rules for propositions do not behave quite like propositions.

By a rule for propositions (briefly, a *rule*) we mean any ordered pair  $(a, x)$  of propositions of the language we are dealing with. A set of rules is thus no more nor less than a binary relation  $R$  over the language, i.e. a set  $R \subseteq L^2$ . Given a set  $X$  of propositions and a set  $R$  of rules, we define the *image* of  $X$  under  $R$ , written as  $R(X)$ , in the standard manner of elementary set theory:  $y \in R(X)$  iff there is an  $x \in$

$X$  with  $(x,y) \in R$ . A set  $X$  is said to be closed under  $R$  iff  $R(X) \subseteq X$ , i.e. iff whenever  $x \in X$  and  $(x,y) \in R$  then  $y \in X$ .

Let  $R \subseteq L^2$  be a set of rules, which intuitively will be playing the role of a set of background ‘inference tickets’ ready for application to any set of premises. Let  $A$  be a set of formulae, and let  $x$  be an individual formula.

We say that  $x$  is a *consequence of  $A$  modulo the rule set  $R$* , and write  $A \vdash_R x$  alias  $x \in Cn_R(A)$  iff  $x$  is in every superset of  $A$  that is closed under both  $Cn$  and the rule set  $R$ . In other words, iff  $x$  is in every set  $X \supseteq A$  such that both  $Cn(X) \subseteq X$  and  $R(X) \subseteq X$ . We call an operation a *pivotal-rule consequence* iff it is identical with  $Cn_R$  for some set  $R$  of rules. Note once again that there is not a unique such relation, but many – one for each value of  $R$ .

It is immediate from the definition that  $Cn \leq Cn_R$  so that pivotal-rule consequence  $Cn_R$  is supraclassical. It is not difficult to verify that it is monotonic and also satisfies cumulative transitivity, so that it is a closure operation. We are thus still in the realm of paraclassical inference.

Like pivotal-assumption, and unlike pivotal-valuation consequence, it can also be verified to be compact. But there is an important property possessed by both pivotal-assumption and pivotal-valuation operations that it lacks, *namely disjunction in the premises* (see section 1.2) For example, if  $R = \{(a,x), (b,x)\}$  then  $x \in Cn_R(a)$  and also  $x \in Cn_R(b)$  but  $x \notin Cn_R(a \vee b)$  because  $Cn_R(a \vee b) = Cn(a \vee b)$ , since the last-mentioned set is vacuously closed under  $R$ , i.e. we have  $R(Cn(a \vee b)) = \emptyset \subseteq Cn(a \vee b)$  since  $a, b \notin Cn(a)$ .

Another important property failed by pivotal-rule consequence, is *contraposition*. This is the principle that whenever  $x \in Cn_R(a)$  then  $\neg a \in Cn_R(\neg x)$ . For a direct counterexample, put  $A = \{a\}$  and  $R = \{(a,x)\}$ . Then  $x \in Cn_R(a)$  but  $\neg a \notin Cn_R(\neg x)$ , because  $Cn_R(\neg x) = Cn(\neg x)$ , since again the last-mentioned set is vacuously closed under  $R$ , i.e. we have  $R(Cn(\neg x)) = \emptyset \subseteq Cn(\neg x)$  since  $a \notin Cn(\neg x)$ .

The relationship between the set of pivotal-assumption consequence operations and the set of pivotal-rule ones is analogous to that between the former and the pivotal-valuation relations studied in section 3. *The pivotal-assumption consequence operations are precisely the pivotal-rule ones that satisfy disjunction in the premises.* For on the one hand, every pivotal-assumption consequence operation is itself a pivotal-rule one – given  $Cn_K$ , simply put  $R = \{(t,k) : k \in K\}$  where  $t$  is a tautology, and it is easily checked that  $Cn_K = Cn_R$ . Conversely, when  $Cn_R$  is a pivotal-rule consequence then as we have shown it is a compact supraclassical closure operation; so when in addition it also satisfies disjunction in the premises, then by the representation theorem of section 2.1 it is also a pivotal-assumption consequence operation.

To end this section, we note a point that will be vital when we pass to default-rule consequence in the next section. It is possible to reformulate the definition of

pivotal-rule consequence in an *inductive* manner. We simply put  $Cn_R(A) = \cup\{A_n : n < \omega\}$  where  $A_1 = A$  and  $A_{n+1} = Cn(A_n \cup R(A_n))$ . The equivalence of this with the original definition is easy to verify using the compactness of classical consequence  $Cn$ .

We can indeed go further and reformulate the inductive definition so that it has only singleton increments in the inductive step. To this, fix an ordering  $\langle R \rangle$  of the set  $R$  of rules - that is, fix a sequence  $(a_i, x_i)_{i < \alpha}$  of all the rules in  $R$ , without repetitions, with  $\alpha$  a positive integer or  $\omega$  according to the cardinality of  $R$  (which we assume to be finite or countable). Then put  $Cn_{\langle R \rangle}(A) = \cup\{A_n : n < \omega\}$  where  $A_1 = A$  as before but now  $A_{n+1} = Cn(A_n \cup \{x\})$ , where  $(a, x)$  is the *first* rule in  $\langle R \rangle$  such that  $a \in A_n$  but  $x \notin A_n$ . In the case that there is no such rule, then we put  $A_{n+1} = Cn(A_n)$ .

To avoid a proliferation of symbols, we are using the same notation  $A_n$  for the terms of this sequence as for the previous sequence, but evidently it is quite a different one, making much smaller jumps with singleton increments. It should be noted that the construction does not imply that rules are applied in the order in which they occur in  $\langle R \rangle$ . When we form  $A_{n+1}$ , the rule  $(a, x)$  may occur earlier in the sequence  $\langle R \rangle$  than some rule  $(b, y)$  that we have already applied, since the passage from  $A_{n-1}$  to  $A_n$  may introduce the body  $a$  for the first time. Also note that by the terms of the definition, once a rule  $(a, x)$  is applied, it is never applied again, since its head  $x$  is in all the subsequent sets  $A_{n+k}$ .

Although the terms  $A_n$  of the second sequence are not the same as those of the first sequence, their union is the same, i.e.  $Cn_{\langle R \rangle}(A) = Cn_R(A)$ . This implies also that the choice of ordering makes no difference to the final result.

These inductive definitions of pivotal-rule consequence may seem like long ways of expressing something that the original definition says quite briefly. And so they are. But we will see in the next section that they - and especially the singleton-increment one - provide the key for passing to the nonmonotonic operations of default-rule consequence.

## 4.2. From Pivotal Rules to Default Rules

The consequence operations  $Cn_R$  defined by using pivotal rules are, as we have seen, monotonic. Once again we pass to nonmonotonicity by allowing the set  $R$  of rules to vary with  $A$ , or more precisely, by requiring the set of those rules in  $R$  that are actually applied in the inductive construction to vary with  $A$ . This is done by imposing consistency constraints on their application.

There are a number of different ways of going about this. We will begin by describing one of the best known, due to Reiter (1980), referred to as his system of 'normal defaults'.

We take as our starting point the inductive definition of pivotal-rule consequence, specifically the one with singleton increments. The new idea is to allow rules to be applied one by one, under the guidance of the ordering  $\langle R \rangle$ , so long as the application does not generate inconsistency.

As before, we fix an ordering  $\langle R \rangle$  of the given set  $R$  of rules. As before, we put  $C_{\langle R \rangle}(A) = \cup\{A_n : n < \omega\}$ , and set  $A_1 = A$ , but the subsequent terms are defined differently. We put  $A_{n+1} = Cn(A_n \cup \{x\})$ , where  $(a, x)$  is the *first* rule in  $\langle R \rangle$  such that  $a \in A_n$  but  $x \notin A_n$  and  $x$  is consistent with  $A_n$ . In the case that there is no such rule, then we put  $A_{n+1} = Cn(A_n)$ .

The introduction of the consistency constraint in the induction step has many consequences. One is that the identity of  $C_{\langle R \rangle}(A)$  will now vary with the particular ordering  $\langle R \rangle$  of  $R$ . For example, if  $A = \{a\}$  and  $R = \{(a, x), (a, \neg x)\}$ , then if we order  $R$  in one way we get  $C_{\langle R \rangle}(A) = Cn(x)$  while if we take the reverse order we get  $C_{\langle R \rangle}(A) = Cn(\neg x)$ . We thus have not one but many sets  $C_{\langle R \rangle}(A)$ , one for each choice of the ordering  $\langle R \rangle$ . Another consequence is that these operations are nonmonotonic. For example, if  $A = \{a\}$  and  $R = \{(a, x)\}$  then as we have seen  $C_{\langle R \rangle}(A) = Cn(x)$ , but when  $A$  is increased to  $B = \{a, \neg x\}$  then  $C_{\langle R \rangle}(B) = Cn(\emptyset)$ .

The sets  $C_{\langle R \rangle}(A)$  for varying orderings  $\langle R \rangle$  of  $R$  coincide with what Reiter (1980) calls *extensions* of  $A$  under the (normal) rules  $R$ . Reiter's original definition was formulated rather differently, but the two are equivalent. Each ordering uniquely determines some extension, and every extension is determined by some ordering.

The operations  $C_{\langle R \rangle}$  are themselves of interest, and if one is equipped with a preferred ordering  $\langle R \rangle$  of  $R$ , one may wish to go no further. But one can also define a consequence  $C_R$  that is independent of any particular order, by intersecting all the  $C_{\langle R \rangle}$ . In other words, we may put  $C_R(A) = \cap\{C_{\langle R \rangle}(A) : \langle R \rangle \text{ an ordering of } R\}$ . In the relational notation:  $A \sim_R x$  iff  $A \sim_{\langle R \rangle} x$  for every possible ordering  $\langle R \rangle$  of the rule-set  $R$ . We call this relation/operation *default-rule consequence*. It coincides with what is usually known as 'consequence using normal Reiter default rules with the sceptical policy on extensions'.

We summarize the principal message of this section. By working with any fixed set of rules we obtain a natural kind of supraclassical consequence relation, which we have called a pivotal-rule consequence. It is monotonic and indeed a closure operation. It differs from the systems of pivotal-assumption and pivotal-valuation consequence, most conspicuously in failing the Horn rule of disjunction of premises. We can also allow the set of rules to vary with the premises of the inference – more precisely, make their step-by-step depend on a consistency constraint that varies with the premise set. In this way, we also obtain quite naturally the well-known normal default consequence operations of Reiter. It is also possible to obtain in essentially the same manner a range of generalizations and variants, for which we refer the reader to the Guide to Further Reading.

## 5. Conclusion

We have sketched three main ways of obtaining more from a set of premises than is classically authorized. They are qualitative in the sense that they do not use probability distributions or similar maps into the real unit interval.

One procedure is to use *background assumptions* that are added to the premises of each inference. Another is to restrict the set of *valuations* of the premises. A third is to deploy additional *rules* alongside classical consequence.

These generate three kinds of *monotonic* relation, which we have called *pivotal-assumption*, *pivotal-valuation* and *pivotal-rule* relations. The first two coincide in the finite case, but can differ in the infinite case; the third is significantly different even in the finite case. Each is of interest in its own right. They are all *paraclassical*, that is, supraclassical closure relations.

Each of these kinds of paraclassical logic serves as an antechamber for the definition of a well-known kind of *nonmonotonic* consequence relation. These are obtained essentially by allowing the background assumptions, valuations, or application of the rules to sway in a systematic way with the premises under consideration. In this way we obtain *default-assumption*, *default-valuation*, and *default-rule* consequences. They coincide with well-known systems of inference using Poole systems, preferential models, and normal Reiter default rules. By varying the constructions, one can also obtain a range of other kinds of nonmonotonic consequence that have been studied in the literature.

The take-home story is in Table 1 below.

**Table 1: Classical, Paraclassical and Nonmonotonic Consequence Operations**

Classical consequence $C_n$		
Use additional assumptions	Restrict the set of valuations	Use additional rules
Pivotal-assumption consequence operations $C_{n_K}$	Pivotal-valuation consequence operations $C_{n_W}$	Pivotal-rule consequence operations $C_{n_R}$
Paraclassical	Paraclassical	Paraclassical
- Disjunction in premises: yes	- Disjunction in premises: yes	- Disjunction in premises: no
- Compact: yes	- Compact: no	- Compact: yes
Vary additional assumptions with the premises	Vary valuation set with the premises	Vary additional rules with the premises

---

- e.g. using consistency constraint	- e.g. using minimalization	- e.g. using consistency constraint
Default-assumption consequence operations $C_K$	Default-valuation consequence operations $C_W$	Default-rule consequence operations $C_R$
(Poole systems and variants)	(preferential systems and variants)	(Reiter systems and variants)

---

## Guide to Further Reading

### General

The picture sketched in these pages is set out in greater detail in Makinson (2003), with proofs of some of the central points and examination of a range of generalizations and variants of all three paths to nonmonotonicity. For an accessible textbook giving a broad coverage of different kinds of nonmonotonic logic, see Antoniou (1997).

### Using Additional Assumptions

The representation theorem for pivotal-assumption consequence appears to have been part of the folklore for decades. It is proven with different terminology in Rott (2001) section 4.3 Observation 5.1, and in its present guise in Makinson (2003).

Default-assumption consequence has a rather long and complicated history. For singleton  $A = \{a\}$  the operation was defined by Alchourrón and Makinson (1982), but as an account of the *revision* of a belief set  $K$  by a new belief  $a$ . The family of all subsets of  $K$  maximally consistent with  $A$  was also studied by Poole (1988), but as part of a study of *abduction*, i.e. the formation of hypotheses to explain data. For an overview, see e.g. Makinson (1994).

### Restricting the Set of Valuations

Schlechta's counterexample may be found in his paper (1992) or in the proof of Observation 3.4.10 in the overview Makinson (1994).

Preferential consequence relations were introduced by Shoham (1988). A classic presentation is Kraus, Lehmann and Magidor (1990). For an overview see for example Makinson (1994).

## Using Additional Rules

The seminal presentation of default rules Reiter (1980), which is still an excellent introduction. For variants and generalizations see for example the overview in Delgrande, Schaub and Jackson (1994) or presentations in the books of Lukaszewicz (1990), Marek and Truszczyński (1993), Antoniou (1997).

## References

Alchourrón, Carlos and David Makinson 1982. 'On the logic of theory change: contraction functions and their associated revision functions', *Theoria* 48: 14-37.

Antoniou, Grigoris 1997. *Nonmonotonic Reasoning*. MIT Press: Cambridge Mass.

Delgrande, J.P., T. Schaub and W.K. Jackson (1994). 'Alternative approaches to default logic' *Artificial Intelligence* 70: 167-237.

Kraus, Lehmann and Magidor 1990. 'Nonmonotonic reasoning, preferential models and cumulative logics', *Artificial Intelligence* 44: 167-207.

Lukaszewicz, W. 1990. *Non-Monotonic Reasoning – Formalization of Commonsense Reasoning*. Ellis Horwood.

Makinson, David 1994. 'General Patterns in Nonmonotonic Reasoning', in *Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3*, ed. Gabbay, Hogger and Robinson. Oxford University Press, pages 35-110.

Makinson, David 2003. 'Bridges between classical and nonmonotonic logic'. To appear in *Journal of the Interest Group in Propositional Logic (IGPL)*.

Marek, V.W. and M. Truszczyński 1993. *Nonmonotonic Logic: Context Dependent Reasoning*. Springer, Berlin

Poole, David 1988. 'A logical framework for default reasoning', *Artificial Intelligence* 36: 27-47.

Rott, Hans 2001. *Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*, Clarendon Press, Oxford UK (Oxford Logic Guides n° 42).

Schlechta, Karl 1992. 'Some results on classical preferential models', *Journal of Logic and Computation* 2: 676-686.

Shoham, Yoav 1988. *Reasoning About Change*. MIT Press, Cambridge USA.

## **Acknowledgements**

Thanks to Björn Bjurling, Michael Freund, Donald Gillies, Daniel Lehmann, Philippe Mongin, Wlodek Rabinowicz, Karl Schlechta and Bernard Walliser for valuable comments on drafts.