"Nonmonotonic" does not mean "probabilistic"

Stenning, K.; van Lambalgen, M.

DOI
10.1017/S0140525X0900048X

Publication date
2009

Document Version
Final published version

Published in
Behavioral and Brain Sciences

Citation for published version (APA):
“Nonmonotonic” does not mean
“probabilistic”

doi:10.1017/S0140525X0900048X

Keith Stenning\textsuperscript{a} and Michiel van Lambalgen\textsuperscript{b}
\textsuperscript{a}Human Communication Research Center, University of Edinburgh, Edinburgh
EH8 9LW, United Kingdom;\textsuperscript{b}Institute for Logic, Language and Computation,
Department of Philosophy, University of Amsterdam, Amsterdam, 1012CP,
The Netherlands.

Keith@inf.ed.ac.uk
www.inf.ed.ac.uk/people/staff/Keith_Stenning
M.vanLambalgen@uva.nl
http://staff.science.uva.nl/~michiel

Abstract: Oaksford & Chater (O&C) advocate Bayesian probability as a
way to deal formally with the pervasive nonmonotonicity of common
sense reasoning. We show that some forms of nonmonotonicity cannot
be treated by Bayesian methods.

One argument that Oaksford & Chater (O&C) proffer in Bayes-
ian Rationality (Oaksford & Chater 2007, henceforth BR) for
a probabilistic approach to cognition is the pervasiveness of non-
monotonicity in reasoning: almost any conclusion can be over-
turned, if additional information is acquired. They claim that
nonmonotonic inferences fall outside the scope of logical
methods, and that probability theory must be preferred.

This judgment of the authors does not reflect the present state of
play (see, e.g., Antoniou 1997, especially Part V). There are
several good formalisms for nonmonotonic reasoning in
existence, some of which even have a computational complexity
that is vastly less than that of classical logic. More importantly,
even in those cases where probabilistic modeling is in principle
useful, one sometimes needs to supplement the model with a
nonmonotonic logic to account for all the data. We illustrate
this here by using an example from Chapter 5 of BR, not men-
tioned in the present Précis, but relevant throughout: the sup-
pression task (Byrne 1989). When subjects are presented with the
modus ponens material “If she has an essay she studies late
in the library. She has an essay,” they almost universally draw
the conclusion: “She studies late in the library.” When instead
they are presented with the same two premises plus the
premise: “If the library is open she studies late in the library,”
about half of them withdraw the inference. Logically speaking,
we here have an instance of nonmonotonicity: addition of a
premise leads to withdrawal of the original conclusion.

O&C argue that their probabilistic model – which represents
the conditional “if p, then q” as a conditional probability
\( P(q|p) \) – can account for suppression effects. To explain
the model and our criticism, we must first state the main Bayesian
mechanism for updating probabilities, “Bayesian conditionaliza-
tion”: \((\text{BaCo})\). The absolute subjective probability of event
\( D \) given that evidence \( E \) has been observed is equal to the con-
ditional probability \( P(D|E) \). Here \( E \) should encompass available
evidence. In this context it is important to note that Bayesian con-
ditionalization is a nonmonotonic principle: An extension of
the evidence \( p \) may invalidate a previous posterior probability for
\( q \) derived by \((\text{BaCo})\) applied to \( P(q|p) \).

Informally, the O&C model works likes this: One is given the
conditional probability \( P(q|p) \) with value, say, 1 – e for some posi-
tive but small \( e \). If few exceptions are salient, \( e \) is small, and, given
\( p \), \((\text{BaCo})\) yields that \( q \) can be concluded with high probability
(namely, 1 – \( e \)). The second conditional highlights a possible
exception (the library being closed), which leads to an increase
in \( e \), and hence to a decrease in the a posteriori propor-
tion of \( q \). But appealing though this picture is, it is not
Bayesian (Sten-
ning & van Lambalgen 2008b). Consider two putative Bayesian
processes that can change the value of \( P(q|p) \) when new possible
exceptions – say, not-\( r \) with probability 1 – \( r \) – become
salient.

(1) The probability space is enlarged from \( \{p, q\} \) to \( \{p, q, r\} \)
(where \( r \) stands for, say, “the library is open”), and this leads to
a new representation of \( P(q|p) \). One may write

\[
\begin{align*}
P(q|p) & = P(q|p \land g) \\
& = \frac{P(q|p \land g \land r) + P(q|p \land g \land \lnot r)}{P(p)} \\
& = \frac{P(q|p \land g \land r)}{P(p)} + \frac{P(q|p \land g \land \lnot r)}{P(p)} \\
& = \frac{P(q|p \land g \land r)}{P(p)} + \frac{P(q|p \land \lnot g \land r)}{P(p)} \\
& = P(q|p \land r)
\end{align*}
\]

Figure 1.

where the last equality follows under the plausible assumption
that \( p \) and \( r \) are independent. O&C assume that the subject
assigns a lower probability to \( P(q|p) \) in the enlarged repre-
sentation \( P(q|p \land r)P(r) \), suggesting that this is because the subject
lowers \( P(r) \) from 1 to a value smaller than 1 when becoming
aware of possible exceptions. In probability theory, \( P(q|p \land r) \)
and \( P(r) \) would simply be a different representation of \( P(q|p) \), and
the values of these expressions must be the same. There are no
rationality principles governing changes in probabilities when
enlarging the probability space, or rather there is one such
probability distribution first assigns probability $r$ upon becoming aware of the second conditional “if $r$ then $q$,” a nonzero probability. The trouble with such a suggestion is that from a Bayesian point of view, the transition from the a priori probability $P(\neg r) = 0$ to the a posteriori $P(\neg r) > 0$ is not allowed, because this cannot be achieved via (BaCo): conditiona-
izing on more evidence cannot make a null probability positive. One thus needs an additional rationality principle (beyond (BaCo)) governing such transitions. In the absence of such a principle, one has to assume that the probabilities of all non-
salient exceptions (such as $\neg r$) are initially very small but nonzero. This increases the computational complexity of prob-
abilistic reasoning enormously: One requires massive storage and intricate computations to maintain consistency of the prob-
ability assignment.

These considerations show that in order to account for the data on the suppression task any probabilistic model needs to be sup-
plemented with a theory about nonmonotonic and non-Bayesian, but still somehow rational, changes in degrees of belief. One may then question whether a probabilistic model is necessary at all; Stenning and van Lambalgen (2005; 2008a) provide a model cast entirely in terms of nonmonotonic logic.