"Nonmonotonic" does not mean "probabilistic"

doi:10.1017/S0140525X0900048X

Keith Stenning[a] and Michiel van Lambalgen[b]

[a]*Human Communication Research Center, University of Edinburgh, Edinburgh EH8 9LW, United Kingdom;* [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/ ~ michiell

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 *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) for a probabilistic approach to cognition is the pervasiveness of nonmonotonicity in reasoning: almost any conclusion can be overturned, 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 logical 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 mentioned in the present *Précis*, but relevant throughout: the suppression 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 conditionalization": (BaCo). The absolute subjective probability of event D given that evidence E has been observed is equal to the conditional probability $P(D|E)$. Here E should encompass available evidence. In this context it is important to note that Bayesian conditionalization is a nonmonotonic principle: An extension of the evidence p may invalidate a previous posterior probability for q derived by (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 positive but small e. If few exceptions are salient, e is small, and, given p, (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 probability of q. But appealing though this picture is, it is *not* Bayesian (Stenning & 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 $1 - P(r)$ – 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

$$P(q \mid p) = \frac{P(p \wedge q)}{P(p)} = \frac{P(p \wedge q \wedge r)}{P(p)} + \frac{P(p \wedge q \wedge \neg r)}{P(p)} = \\ = \frac{P(p \wedge q \wedge r)}{P(p)} = \frac{P(p \wedge q \wedge r)}{P(p \wedge r)} \frac{P(p \wedge r)}{P(p)} = P(q \mid p \wedge r) P(r)$$

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 representation $P(q|p \,\&\, r)P(r)$, 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, $P(q|p \,\&\, r)$ $P(r)$ 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

principle, that $P(q|p)$ remains the same when computed on an enlarged space. This is the only way in which one can guarantee that enlargements of the probability space in the limit lead to a coherent probability distribution – the starting point of Bayesian rationality.

(2) An orthodox Bayesian alternative would be a construction in which the probability spaces remain the same (namely, the universal space based on all possible propositions), but the probability distributions change. In our toy world, the probability space is in both cases $\{p, q, r\}$, but one could assume that the probability distribution first assigns probability *0* to *not-r*, and, 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 $P(not\text{-}r)=0$ to the a posteriori $P(not\text{-}r) > 0$ is not allowed, because this cannot be achieved via (BaCo): conditionalizing 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 nonsalient exceptions (such as *not-r*) are initially very small but nonzero. This increases the computational complexity of probabilistic reasoning enormously: One requires massive storage and intricate computations to maintain consistency of the probability assignment.

These considerations show that in order to account for the data on the suppression task any probabilistic model needs to be supplemented 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.