A Natural Generalization of Jeffrey Conditioning

The problem that Wagner considers in his paper "Generalized Probability Kinematics" is that of how to revise one's probabilities for some set of variables upon receiving information best encoded as a probability distribution over another set of variables. His solution consist in effect of looking at constraints that the new information might plausibly be said to place on the probabilities of interest and then adopting a 'conservative' rule of taking the new evidence to be uninformative about any possibilities not covered by these constraints. Unlike maxent the adoption of this rule is taken to be based on an explicit judgement on the part of the agent as to its appropriateness in the particular case and is not to be applied mechanically.

The author's main complaint is that Wagner's approach is 'anti-Bayesian' because it does proceed by defining a prior on the joint probability space containing both the original set of variables and those representing the new information. He or she shows that if one proceeds in the 'correct' way, namely by using maxent to define a prior distribution over the joint space and then conditionalise on it, the result is exactly the solution recommended by Wagner. This is a nice observation, but does not do a great deal to address the misgivings typically expressed about maxent and in particular the paradoxes that seem to be generated by applying it mechanically.

Let's look at the author's simplified Linguist problem to illustrate what is at stake. We assume prior probabilities P over the space {Catholic, Protestant} of {0.5;0.5} and suppose that information is received that is best encoded by a probability distribution {0.6;0.4} over events {w1,w2} where w1 implies Catholic but w2 is consistent with both possibilities. The constraints on the agent's posterior probabilities Q can then be represented by:

	W1	**W2**	**Q**
Catholic	0.6	?	?
Protestant	0	?	?
Q	0.6	0.4	1

Wagner's solution consists in taking W2 to be uninformative on the issue of Catholic versus Protestant, formalised as the condition that the ratio of the conditional probabilities of each given W2 is equal to the ratio of their prior probabilities. This gives:

	W1	**W2**	**Q**
Catholic	0.6	0.2	0.8
Protestant	0	0.2	0.2
Q	0.6	0.4	1

The author notes that the same solution can be obtained by defining a prior probability space for the problem using maxent, which he/she claims implies:

	W1	**W2**	**P**
Catholic	0.3	0.2	0.5
Protestant	0.3	0.2	0.5
P	0.6	0.4	1

Then given that P(Protestant|w1) = 0, conditionalising on the information received yields the second table.

But this recovery of the Wagner solution is misleading because it reconstructs the revision inducing experience as one of learning for certain that Catholic, against the backdrop of a particular prior on the joint space. Where does this prior come from? Not by applying maxent to the agent's initial information state (when nothing is known about w1 and w2), for that would yield:

	W1	**W2**	**P**
Catholic	0.25	0.25	0.5
Protestant	0.25	0.25	0.5
P	0.5	0.5	1

And conditionalising on ¬(Protestant and w1) would then yield a different solution to Wagner's.

Instead the prior is generated by taking the newly acquired probabilities of w1 and w2 as their *prior* probabilities. But if this were correct then no revision at all should be required. The reason why a revision is in fact induced is because the author treats the learning experience as one in which the inconsistency of w1 and Protestant is learnt. But this clearly not at all the kind of experience experience which Wagner is concerned with. So although the author is correct that Wagner's solution can be recovered using a maxent rule, it remains entirely opaque as to why using maxent in this way (i.e. by applying it subject to the constraints imposed by the newly acquired probabilities) is appropriate.

The underlying issue is a broader one. Both Wagner's conservative rule and maxent are justified by the thought that one should revise no more than is required by the evidence. But there is a crucial difference. Wagner considers the application of the conservative rule to be grounded in a subjective judgement of appropriateness to the context. He does not endorse it as a general procedure. Maxent on the other hand is proposed as a general rule, to be applied to any circumstance. But what the author's own reconstruction shows is that it cannot be applied mechanically: choosing the constraints relative to which entropy is to be maximised is crucial. Critics of maxent often complain that there is no principled method for choosing these constraints. This paper does little to show that this complain is unjustified.