Comments on Carl Wagner's *Jeffrey Conditioning and External Bayesianity*

Stephen Petersen
steve@stevepetersen.net

Department of Philosophy
Niagara University

Formal Epistemology Workshop 2008
University of Wisconsin-Madison

Outline

Jeffrey conditioning

- Jeffrey conditioning allows updating in Bayesian style when the evidence is uncertain.
- A weighted average, essentially, over classically updating on the alternatives.
- Unlike classical Bayesian conditioning, this allows learning to be unlearned.

Jeffrey conditioning and commutativity

- Field 1978: Jeffrey conditioning needs an "input factor" to measure the change in the events directly affected by learning.
- He proposes, in effect, Bayes factors; for $A, B \in \mathscr{P}(\Omega)$,

$$\beta_{q,p}(A : B) = \frac{q(A)/q(B)}{p(A)/p(B)}$$

- He shows that this paramaterization preserves commutativity.
- (unlike measuring learning by the posterior evidential probabilities)

Wagner's argument from mathematical elegance

Wagner's *Uniformity Rule*

Bayesians should represent "identical learning" by sameness of Bayes factors across atomic events.

- Wagner argues for this rule with a pile of mathematical elegance.
- Today he showed how it can capture commutativity of pooling operators.
- Elsewhere he extends Field's result to infinite sample spaces with countable partitions.

Wagner's argument from mathematical elegance

- He also shows the same trick of preserving Bayes factors—when applied to *conditional* rather than evidential probabilities—can generalize Jeffrey's solution to the historical old evidence problem for uncertain updating.
- Along the way he shows why rival representations of learning (relevance quotients, probability differences) can't do the same as neatly.
- Finally, it has a nice tie with a recent plausible metric from Chan & Darwiche for probability measures over a finite sample space.

Chan-Darwiche distance and the uniformity rule

$$
\begin{aligned}
\mathrm{CD}(p,q) &= \log \max_{\omega\in\Omega} \frac{q(\omega)}{p(\omega)} - \log \min_{\omega\in\Omega} \frac{q(\omega)}{p(\omega)} \\
&= \log \frac{\max_{\omega\in\Omega} q(\omega)/p(\omega)}{\min_{\omega'\in\Omega} q(\omega')/p(\omega')} \\
&= \max_{\omega,\omega'\in\Omega} \log \frac{q(\omega)/p(\omega)}{q(\omega')/p(\omega')} \\
&= \max_{\omega,\omega'\in\Omega} \log \frac{q(\omega)/q(\omega')}{p(\omega)/p(\omega')} \\
&= \max_{\omega,\omega'\in\Omega} \log \beta_{q,p}(\{\omega\} : \{\omega'\}) \\
&= \max_{A,B\in\mathscr{P}(\Omega)-\emptyset} \log \beta_{q,p}(A : B)
\end{aligned}
$$

Problems with "identical learning"

- As Wagner is aware, this does not settle philosophical questions about "identical learning".
- There are a number of cases that seem to show this is still a messy notion.

Identical learning and sensory experience

- Garber: Bayes factors can't capture learning in the sense of *sensory experience*.
- Otherwise, repeating an uncertain sense experience will, by repeated applications of Bayes factors, drive you toward certainty.
- Wagner: we should therefore divorce identical learning from sense experiences; "we learn nothing new from repeated glances and so all Bayes factors beyond the first are equal to one."

Döring's case

- It sometimes seems very odd, though, to divorce learning from sensory experience.
- Döring's case:
  - You have a low prior some shirt is blue, I have a high one.
  - We catch an identical glimpse under a neon light, and it looks blue-green.
  - Your posterior should be higher, mine lower.
  - Therefore our Bayes factors differ.
  - Therefore we didn't learn the same thing.
- In some sense maybe this is right—but in some important sense we surely *did* learn the same thing.

Factoring out priors

- Field wanted to factor out the priors with his "input factor".
- Döring's case shows, though, that priors make a difference to whether you undergo "identical learning" in this sense.
- Field seemed to hope that factoring out priors would thereby capture just the new sensory experience.
- But there are a few non-equivalent ways to factor out a starting point for probability movement, depending on your purpose.
  - measure only how far you move
  - measure only the pushing force
  - measure only where you end up

Priors-relative learning

- In other cases it's plausible learning depends on the priors.
- Skyrms case (in Lange paper):
  - I catch a dim fleeting glimpse of a crow.
  - I thus assign it a relatively low probability of being black.
  - I update on this uncertainty,
  - and thereby disconfirm my hypothesis that all crows are black.
- "I could disconfirm lots of theories just by running around at night."

Lange's take on Skyrms

- Lange: if "the raven looks about the way that any dusky colored object would be expected to look under those conditions,"
- then we should perhaps instead think of this sensory experience as *inflating* the prior odds that this crow is black—only more *slightly* than usual.
- Thus the Wagner-Field uniformity rule looks appropriate.
- Lange's suggestion: "...two agents are undergoing the same sensory experience exactly when it is the case that *had* the two agents begun with the *same* prior probability distribution, then they *would* as a result of their actual sensory experiences have imposed exactly the *same* constraints on that distribution, ... no matter what the two agents' common prior probability distribution had been."

Osherson

- Similarly, Osherson:
  - If one glimpse of clouds moves my subjective probability of rain from .3 to .7,
  - and (in a scenario with alternate priors) the glimpse of clouds moves my subjective probability from .5 to .7,
  - then they must have been different sensory experiences.
- This seems to suggest sensory experience should be determined by something like Bayes factors.
- (So the Garber case actually involves different sensory experiences?!)

Rosencrantz and commutativity

- It's also not obvious that commutativity should be preserved when updating on uncertain evidence.
- Rosencrantz case (in Lange): "Consider a child who has just knocked over a jar of paint and is wondering whether he is going to get spanked. In one scenario, a parental scowl is followed by good natured laughing, while, in the other, these responses occur in the opposite sequence!"
- Lange:
  - This is classical conditioning, so it will commute.
  - It appears not to because they are *not* the same pieces of evidence in a different order.
  - One is a scowl-into-laugh, another a laugh-into-scowl.

Rosencrantz and commuatitivity

- Lange's response seems too quick to me.
- First, this could easily be a case of Jeffrey conditioning—the expressions could be uncertain evidence for the parent's anger, on which the spanking probability is really updated.
- Given that a video of one transformation could be the reverse of the other, then they can be seen as the same *sensory* experiences in a different order.
- (Think of the frames of the video at 30+ frames per second.)
- The motivation for calling them "different" seems simply to be that they nudge the posterior for spanking in different directions.
- We could admit them as different elements in the sample space, and do classical conditioning—but how plausible is that?

More on commutativity

- Other times it seems clear we want commutativity.
- Case based on Jeffrey (unpublished):
  - You have a tumor that may be malignant, and treat this as uncertain evidence for the claim you will live at least five more years.
  - Histopathologist: .8 probability malignant.
  - Radiologist: .6 probability malignant.
  - Posterior shouldn't depend on the order in which you visit them.
- Döring:
  - Explosion occurs in one of four quadrants of an airplane.
  - You find an intact chunk of the back right.
  - You find an intact chunk of the back left.
  - Posteriors for right *vs.* left should not depend on the order.
- In both these cases, sure looks like cheating to say that it's a different piece of evidence when it happens in a different order.

Concluding hunches

- Maybe sometimes order matters, and sometimes it doesn't.
- Maybe sometimes a sense experience washes out the prior, and sometimes it doesn't.
- Maybe sometimes "same learning" means "same evidential posteriors", and sometimes it means "same evidential Bayes factors".
- Total hunches:
  - The problem is in the variability in specifying the sample space.
  - The attendant *ad hoc*kery will haunt us until we can revive some protocol-sentence-like notion of observation, independent of background theory.
  - Such a notion cannot be revived.