SCORING RULES, UPDATING AND COHERENCE

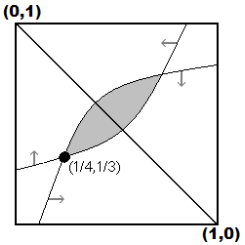
One or Two Thoughts about Accuracy, Epistemic Utility and Coherence

Jim Joyce
Department of Philosophy
The University of Michigan
jjoyce@umich.edu

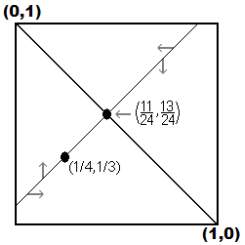
CREDENCE FUNCTIONS AND SCORING RULES

A *credence function* $\boldsymbol{b}$ assigns degrees of belief to propositions in a set $\mathcal{X} = \{X_1, \ldots, X_N\}$.

Given a credence function $\boldsymbol{b}$ and a truth-value assignment $\boldsymbol{v}$ for $\mathcal{X}$, an *epistemic scoring rule* $\boldsymbol{S}$ produces a real number $\boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v}) \geq 0$ that measures the *epistemic disutility* of holding the credences $\boldsymbol{b}$ when $\boldsymbol{v}$ is actual. A perfect score is $\boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v}) = 0$. $\boldsymbol{S}(\boldsymbol{c}, \boldsymbol{v}) > \boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v})$ means that, in terms of epistemic quality, the credences in $\boldsymbol{b}$ are better than those in $\boldsymbol{c}$ in world $\boldsymbol{v}$.



$\text{Brier}(\boldsymbol{b}, \boldsymbol{v}) = \sum_n (b_n - v_n)^2$
$\text{Brier}((p, q), \langle 1,0\rangle) = (1-p)^2 + q^2$
$\text{Brier}((p, q), \langle 0,1\rangle) = p^2 + (1-q)^2$



$\text{Abs}(\boldsymbol{b}, \boldsymbol{v}) = \sum_n |b_n - v_n|$
$\text{Abs}((p, q), \langle 1,0\rangle) = (1-p) + q$
$\text{Abs}((p, q), \langle 0,1\rangle) = p + (1-q)$

INADMISSIBILITY

DEFINITION: $\boldsymbol{b}$ is *inadmissible* relative to $\boldsymbol{S}$ when there is some alternative $\boldsymbol{b}_{\boldsymbol{S}}$ such that $\boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v}) > \boldsymbol{S}(\boldsymbol{b}_{\boldsymbol{S}}, \boldsymbol{v})$ for all $\boldsymbol{v}$. $\boldsymbol{b}_{\boldsymbol{S}}$ will typically depend on $\boldsymbol{S}$.

If $\boldsymbol{S}$ is the Brier score, then $\boldsymbol{b} = (1/4, 1/3)$ is inadmissible because for $\boldsymbol{b}_{\boldsymbol{S}} = (½, ½)$ we have

$$\boldsymbol{B}(\boldsymbol{b}, \langle 1,0\rangle) = 0.337 > \boldsymbol{B}(\boldsymbol{b}_{\boldsymbol{S}}, \langle 1,0\rangle) = 0.25$$
$$\boldsymbol{B}(\boldsymbol{b}, \langle 0,1\rangle) = 0.253 > \boldsymbol{B}(\boldsymbol{b}_{\boldsymbol{S}}, \langle 0,1\rangle) = 0.25$$

Note: Here a coherent credence dominates an incoherent credence.

For the absolute value score, $\boldsymbol{b} = (1/3, 1/3, 1/3)$ is inadmissible. For $\boldsymbol{b}_{\boldsymbol{S}} = (0, 0, 0)$ we have

$$\boldsymbol{A}(\boldsymbol{b}, \langle 1,0,0\rangle) = 4/9 > \boldsymbol{A}(\boldsymbol{b}_{\boldsymbol{S}}, \langle 1,0,0\rangle) = 1/3$$
$$\boldsymbol{A}(\boldsymbol{b}, \langle 0,1,0\rangle) = 4/9 > \boldsymbol{A}(\boldsymbol{b}_{\boldsymbol{S}}, \langle 0,1,0\rangle) = 1/3$$
$$\boldsymbol{A}(\boldsymbol{b}, \langle 0,0,1\rangle) = 4/9 > \boldsymbol{A}(\boldsymbol{b}_{\boldsymbol{S}}, \langle 0,0,1\rangle) = 1/3$$

Note: Here an **inconsistent** credence dominates a coherent credence.

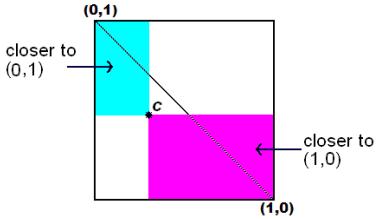
THINGS TO KEEP IN MIND

- The choice of an epistemic scoring rule should reflect our considered views about what sorts of traits make beliefs worth holding (in the context of assessment).
  - Are beliefs being assessed in terms of their value as guides for action?
  - Are we only concerned with their accuracy as representations of the world?
- Different features of scoring rules reflect different epistemic values.
  - E.g., some rules "give points" for epistemic conservatism – they say that if you and I have different credences with identical $\boldsymbol{S}$-scores, then we both can do better by moving our credences closer together (thus farther from the extremes).
  - Others encourage dogmatism, even at the cost of inconsistency. (Absolute Value)
- It is legitimate methodology to reject a putative scoring rule $\boldsymbol{S}$ if $\boldsymbol{c}$ $\boldsymbol{S}$-dominates $\boldsymbol{b}$ when $\boldsymbol{b}$ is superior by the standards of value relevant in the context.
  - E.g., if we value logical consistency more than proposition-by-proposition distance from the truth we should reject the Absolute Value $\boldsymbol{S}$core.

- In contrast with interpretations of scoring rules prevalent in economics, we should ***not*** think of $\boldsymbol{S}$ as something a believer *aims* to maximize by choosing her opinions.
  - Rather, it provides a standard of assessment that we use to evaluate the beliefs of others, and sometimes ourselves, from a kind of third-person perspective. (Joyce 1998) is unclear/confused on this point.
- One should not see the epistemic interpretation of scoring rules as conflicting with interpretations in which the scoring rules measure something the believer values.
- This means that the arguments for coherence presented here do ***not*** compete, and do ***not*** conflict with, more familiar scoring-rule based arguments for coherence: they are two roads to the same destination.
  - One difference: I need to argue for the normative appropriateness of scoring rules, whereas proponents of, say, prevision-based arguments for coherence can assume that the believer has a vested interest in minimizing her penalty.
  - This makes my argument different from deFinetti's.

Some Potentially Desirable Properties of Scoring Rules

Truth-Directedness. If all $\boldsymbol{b}$'s values are closer than $\boldsymbol{c}$'s are to $\boldsymbol{v}$, then $\boldsymbol{S}(\boldsymbol{c}, \boldsymbol{v}) > \boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v})$.

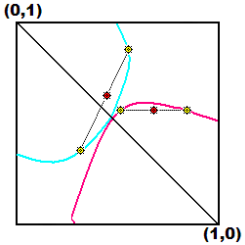


Note: This rules out **calibration** as an acceptable epistemic scoring rule.

Extensionality. $\boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v})$ is a function exclusively of the credences $\langle b_1,\ldots, b_N\rangle$ and the truth-values in $\boldsymbol{v} = \langle v_1,\ldots, v_N\rangle$.

- Nothing about the *content* of propositions in $\mathcal{X}$ matters (e.g., their informativeness or verisimilitude or objective chance).

Convexity. $\frac{1}{2}\cdot\boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v}) + \frac{1}{2}\cdot\boldsymbol{S}(\boldsymbol{c}, \boldsymbol{v}) > \boldsymbol{S}(\frac{1}{2}\cdot\boldsymbol{b} + \frac{1}{2}\cdot\boldsymbol{c}, \boldsymbol{v})$.



- This enforces a kind of conservativism. If you and I score equally well according to $\boldsymbol{S}$, then we can each improve our score by moving closer to one another.

Additivity. $\boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v}) = \Sigma_n \lambda_n\cdot\boldsymbol{s}_n(b_n, v_n)$, where

- each *component function* $\boldsymbol{s}_n$ measures the epistemic utility of $\boldsymbol{b}(X_n) = b_n$ on a scale that decreases/increases in first coordinate when the second coordinate is one/zero
- the w*eights* λ_n are non-negative real numbers summing to one that reflect the degree to which the utilities of credences for X_n matter to overall epistemic utility.

Proper Scoring Rules, Information and Divergence

Propriety. Each coherent credence function must uniquely maximize expected epistemic utility when expectations are taken relative to itself. So, for any partition $\mathcal{X} = \{X_1,..., X_n\}$ if $\boldsymbol{v}^j$ is the truth-value assignment that makes X_j true, then for any *coherent* $\boldsymbol{b}$ and any $\boldsymbol{c}$

$$Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{c})) = \Sigma_j\, b_j\cdot\boldsymbol{S}(\boldsymbol{c}, \boldsymbol{v}^j) > \Sigma_j\, b_j\cdot\boldsymbol{S}(\boldsymbol{b}, \boldsymbol{v}^j) = Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{b}))$$

Each coherent credence function is "immodest" –– it says of itself that it does better than any other credence function in exemplifying the epistemic virtues encoded in $\boldsymbol{S}$.

Some Things to like about Prorpiety

- *De Finetti and Savage*. Penalizing people using proper scoring rules encourages them to reveal their actual credences in "belief elicitation" experiments.
- *Gibbard*. If we are going to use our credences for "guidance" we must expect them to be superior to any other credences.
- *Savage*. The component functions $\boldsymbol{s}(b, 1)$ and $\boldsymbol{s}(b, 0)$ define a strictly proper, additive, extensional rule iff there is some twice differentiable positive function $\boldsymbol{g}$ on [0, 1] with $\boldsymbol{g}'' < 0$ on (0, 1) and $\boldsymbol{s}(b, v) = \boldsymbol{g}(b) + (v - b)\cdot\boldsymbol{g}'(b)$.

INFORMATION AND DIVERGENCE

- Each proper scoring rule $\boldsymbol{S}$ has an associated ***"information measure"***

$$\mathscr{I}_S(\boldsymbol{b}) = Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{b}))$$

- Each proper rule has an associated ***"divergence function"***

$$\mathscr{D}_S(\boldsymbol{c}, \boldsymbol{b}) = Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{c})) - Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{b}))$$

  - $\mathscr{D}_S(\boldsymbol{c}, \boldsymbol{b})$ need **not** be a *metric*, e.g., it often will not be true that $\mathscr{D}_S(\boldsymbol{c}, \boldsymbol{b}) = \mathscr{D}_S(\boldsymbol{b}, \boldsymbol{c})$.

Interpretation:

- $Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{c}))$ gives the degree to which $\boldsymbol{c}$'s credences realize the epistemic virtues encoded in $\boldsymbol{S}$ *on the assumption that* $\boldsymbol{b}$ *best realizes those virtues*.
- $\mathscr{I}_S(\boldsymbol{b})$ is the maximum degree to which $\boldsymbol{S}$'s epistemic virtues can be realized *on the assumption that* $\boldsymbol{b}$ *best realizes those virtues*.
- $\mathscr{D}_S(\boldsymbol{c}, \boldsymbol{b})$ measures the amount by which $\boldsymbol{c}$ falls short of best realizing $\boldsymbol{S}$'s epistemic virtues *on the assumption that* $\boldsymbol{b}$ *best realizes those virtues*.

THE QUADRATIC "PACKAGE"

Brier Score: $\boldsymbol{B}(\boldsymbol{v}, \boldsymbol{b}) = \sum_n (v_n - b_n)^2$

Brier Information: $\mathscr{I}_B(\boldsymbol{b}) = 1 - \sum_n b_n^2$

Note: When $\mathcal{X}$ is a partition this is $\boldsymbol{b}$'s *variance* for the truth-value assignment.

Brier Divergence: $\mathscr{D}_B(\boldsymbol{c}, \boldsymbol{b}) = \sum_n (b_n - c_n)^2$

Note: This is symmetric and is the squared Euclidean distance between $\boldsymbol{c}$ and $\boldsymbol{b}$.

Caution: Euclidean distance is ***not*** a proper scoring rule, and thus not a good way to measure "distances" between probability functions for purposes of comparing their epistemic merits.

THE LOGARITHMIC "PACKAGE"

Log Score: $\boldsymbol{L}(\boldsymbol{v}, \boldsymbol{b}) = \sum_n - ln(|1 - v_n - b_n|)$

Shannon Entropy: $\mathscr{I}_L(\boldsymbol{b}) = -\sum_n b_n \cdot ln(b_n)$

Note: This is the negative of Shannon information.

Kullback-Leibler Divergence: $\mathscr{D}_B(\boldsymbol{c}, \boldsymbol{b}) = \sum_n b_n \cdot ln(b_n/c_n)$

Note: This is not symmetric and it goes to infinity when $\boldsymbol{c}$ is certain of something that $\boldsymbol{b}$ regards as having some positive probability of occurring.

"MECHANICAL" UPDATING

Learning experiences impose constraints on the form of posterior credences by confining the posterior $\boldsymbol{c}$ to some (convex) subset of probabilities $\boldsymbol{C}$.

FACT: Under broad conditions, for any proper scoring rule $\boldsymbol{S}$ and any prior probability $\boldsymbol{b}$ there will be a unique posterior $\boldsymbol{b}_{\boldsymbol{C}}$ among those that satisfy the constraint which minimizes $\boldsymbol{S}$-divergence from $\boldsymbol{b}$, so that $\mathscr{D}_B(\boldsymbol{c}, \boldsymbol{b}) > \mathscr{D}_B(\boldsymbol{b}_{\boldsymbol{C}}, \boldsymbol{b})$ for all $\boldsymbol{c} \neq \boldsymbol{b}_{\boldsymbol{C}}$ in $\boldsymbol{C}$.

- Interpretation: Among the credence functions that satisfy the constraint $\boldsymbol{b}_{\boldsymbol{C}}$ best realizes the epistemic virtues encoded in *S on the assumption that* $\boldsymbol{b}$ *best realizes those virtues tout court*.

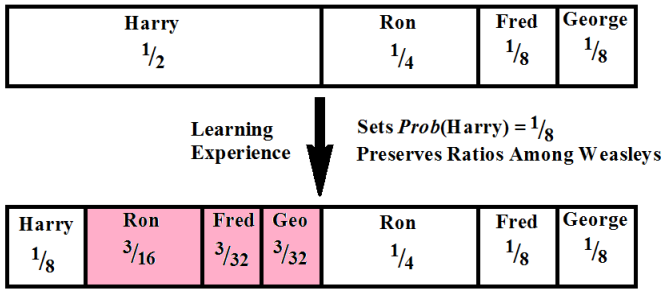
- *Example* (***S****-conditioning*): The new evidence is $\boldsymbol{c}(X) = 1$. $\boldsymbol{b}_{\boldsymbol{C}}$ is the credence function that both makes $\boldsymbol{X}$ certain and, by $\boldsymbol{b}$'s lights, best realizes the $\boldsymbol{S}$-virtues.

- *Example* (***S****-kinematics*): The new evidence comes as a "Jeffrey experience" which involves a direct shift in probabilities $b_n \rightarrow c_n$ over the propositions in some partition $\mathcal{X} = \{X_1, \ldots, X_N\}$. $\boldsymbol{b}_{\boldsymbol{C}}$ is the credence function that among those for which $\boldsymbol{c}(X_n) = c_n$ that, by $\boldsymbol{b}$'s lights, best realizes the $\boldsymbol{S}$-virtues.

A Misuse of This Apparatus

Richard Pettigrew and Hannes Leitgeb have argued that a policy of measuring epistemic accuracy using the Brier score justifies a belief update rule that contradicts ordinary Bayesian conditioning and Jeffrey conditioning (a.k.a. "probability kinematics").

Example of **JC**: You know that Harry Potter or one of three Weasely boys -- Ron, Fred or George -- is planning to meet you. Your priors for these hypotheses are (½, ¼, $^1/_8$, $^1/_8$). You see someone crossing the heath in the dusk. His hair seems reddish. This causes to lower your probability for Harry to $^1/_8$. JC tells you to reapportion your probabilities so that the ratios of probabilities among the Weasleys stays the same.



Formally, if the evidence mandates $\boldsymbol{c}(X_n) = c_n$ for $n = 1, 2,..., N$, then Jeffrey kinematics is characterized by any of the following equivalent conditions:

JC1 $\boldsymbol{c}(\bullet) = \Sigma_n c_n \cdot \boldsymbol{b}(\bullet \mid X_n)$

JC2 "Rigidity": For any $n = 1, 2,..., N$, $\boldsymbol{c}(X_n) = c_n$ and $\boldsymbol{b}(\bullet \mid X_n) = \boldsymbol{c}(\bullet \mid X_n)$.

JC3 "Ratio Invariance": For any $n = 1, 2,..., N$, $\boldsymbol{c}(X_n) = c_n$ and ratios of credences for events that entail X_n do not change, so that $\boldsymbol{b}(A \,\&\, X_n)/\boldsymbol{b}(B \,\&\, X_n) = \boldsymbol{c}(A \,\&\, X_n)/\boldsymbol{c}(B \,\&\, X_n)$ for all propositions A and B (for which the identity makes sense).

Note: These last two characterizations are often seen a conveying the idea that the *only* thing that one learns from the experience is $\boldsymbol{c}(X_n) = c_n$ for all $n = 1, 2,..., N$.

An Interesting Fact about Jeffrey Kinematics

FACT (Diaconis, Zabell): If we measure epistemic utility using the **log score** $\boldsymbol{L}(\boldsymbol{v}, \boldsymbol{b})$, and so use $\mathfrak{D}_L(\boldsymbol{c}, \boldsymbol{b})$ to measure the divergence of a posterior $\boldsymbol{c}$ from the prior $\boldsymbol{b}$, then Jeffrey kinematics gives the unique update probability that

(a) satisfies the constraints $\boldsymbol{c}(X_1) = c_1$, $\boldsymbol{c}(X_2) = c_2$,... $\boldsymbol{c}(X_n) = c_N$

(b) minimizes $\mathfrak{D}_L(\boldsymbol{c}, \boldsymbol{b})$.

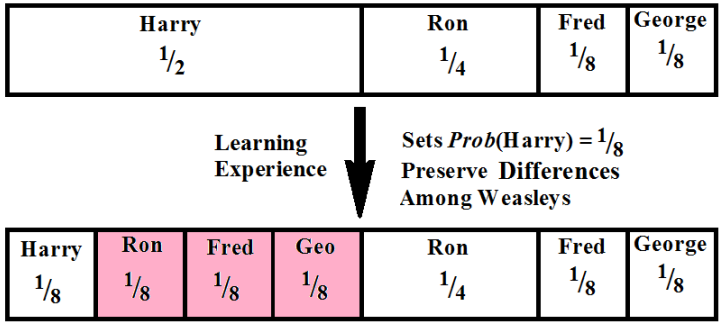
In other words, among probabilities consistent with the experience, JC minimizes Kullback-Leibler divergence from the prior!

Many people (myself included) have portrayed this as an argument for JC. It seems to say that JC involves (a) accepting the new evidence and then (b) adopting the posterior that adds the least amount of new information beyond that already contained in the prior. That seems like a good thing!

NOTE: Since ordinary Bayesian conditioning is a special case of JC, this result applies when the constraint is $\boldsymbol{c}(X) = 1$ as well.

Pettigrew/Leitgeb Kinematics

Example of **PL**: Same story, same priors, same experience but you to reapportion your probabilities so that the ***differences*** of probabilities among the Weasleys stays the same.



Formally, if the evidence mandates $\boldsymbol{c}(X_n) = c_n$ for $n = 1, 2, ..., N$, then PL-kinematics is characterized as follows:

- Assume that each X_n can be uniquely partitioned into a finite set of "worlds", so that $X_n = \omega_{n1} \vee \omega_{n2} \vee \ldots \vee \omega_{nM(n)}$.
- If $\boldsymbol{c}_n > \boldsymbol{b}(X_n)$, then set $\boldsymbol{c}(\omega_{nm}) = \boldsymbol{b}(\omega_{nm}) + [\boldsymbol{c}_n - \boldsymbol{b}(X_n)]/M(n)$ for all $m \leq M(n)$.
- If $\boldsymbol{c}_n < \boldsymbol{b}(X_n)$, there is a unique number $d_n < 0$ such that $\Sigma_n \max\{0, \boldsymbol{b}(\omega_{nm}) + d_n\} = c_n$, and set $\boldsymbol{c}(\omega_{nm}) = \max\{0, \boldsymbol{b}(\omega_{nm}) + d_n\}$.

  (If this seems opaque to you, don't sweat it – I am going to encourage you to forget PL-kinematics as fast as possible after the talk.)

  Note: PL-kinematics is inconsistent with ordinary Bayesian conditioning because, e.g., it can raise the probability of cells of the partition from zero to some positive number.

AN INTERESTING FACT ABOUT PETTIGREW LEITGEB KINEMATICS

FACT (Pettigrew, Leitgeb): If we measure epistemic utility with the **Brier score** $\boldsymbol{B}(\boldsymbol{v}, \boldsymbol{b})$, and use $\mathscr{D}_{\boldsymbol{B}}(\boldsymbol{c}, \boldsymbol{b})$ to measure the divergence of a posterior probability $\boldsymbol{c}$ from a prior $\boldsymbol{b}$, then PL-kinematics gives the unique update probability that

(a) satisfies the constraints $\boldsymbol{c}(X_1) = c_1$, $\boldsymbol{c}(X_2) = c_2, \ldots \boldsymbol{c}(X_n) = c_N$

(b) minimizes $\mathscr{D}_{\boldsymbol{B}}(\boldsymbol{c}, \boldsymbol{b})$.

In other words, among probabilities consistent with the experience, PL uniquely minimizes Brier divergence from the prior.

Pettigrew and Leitgeb portray this as an argument for their conditioning rule! It seems to say that PL involves only (a) adding the new evidence and (b) selecting the posterior consistent with that evidence that best realizes the epistemic virtues encoded in the Briar score. That seems like a good thing!

A FORCED CHOICE?

Given the FACTS, it looks like we are forced to ***choose*** between the logarithmic package of $\boldsymbol{L}(\boldsymbol{v}, \boldsymbol{b})$, $\mathscr{D}_{\boldsymbol{L}}(\boldsymbol{c}, \boldsymbol{b})$, $\mathscr{I}_{\boldsymbol{L}}(\boldsymbol{b})$ and Jeffrey Kinematics, and the quadratic package of $\boldsymbol{B}(\boldsymbol{v}, \boldsymbol{b})$, $\mathscr{D}_{\boldsymbol{B}}(\boldsymbol{c}, \boldsymbol{b})$, $\mathscr{I}_{\boldsymbol{B}}(\boldsymbol{b})$ and PL-Kinematics.

- If you like JC and ordinary conditioning and if you dislike PL (and there's a lot to dislike), you must embrace ***Log*** and reject ***Brier***.
- If you like ***Brier*** (and there's a lot to like), you must embrace PL and reject JC and ordinary conditioning.

Even worse, for other proper scoring rules there will (I'm pretty sure) be other update rules that uniquely minimize divergence! What are we supposed to say about them?

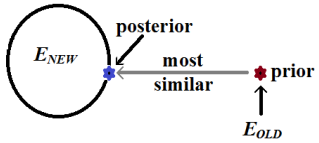
A FALSE CHOICE!

The whole mechanical updating picture is wrong-headed: it is illegitimate to update by maximizing prior epistemic utility subject to a constraint.

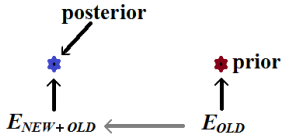
- Recall that $Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{c}))$ reflects the degree to which $\boldsymbol{c}$'s credences realize the epistemic virtues encoded in $\boldsymbol{S}$ *on the assumption that* $\boldsymbol{b}$ *best realizes those virtues*.
- $\mathscr{D}_{\boldsymbol{S}}(\boldsymbol{c}, \boldsymbol{b})$ is the amount by which $\boldsymbol{c}$ falls short of best realizing $\boldsymbol{S}$'s epistemic virtues *on the assumption that* $\boldsymbol{b}$ *best realizes those virtues*.
- But, insofar as the experience imparts *new evidence* (e.g., that it's 7:1 that a Weasley is on the way), ***b* loses** its status as the credence function that best realizes the $\boldsymbol{S}$-virtues. After all, the evidence shows that $\boldsymbol{b}$ is wrong about the probabilities of the X_n (e.g., it's wrong about how likely a Weasley is to come).
- So, knowing $\mathscr{D}_{\boldsymbol{S}}(\boldsymbol{c}^*, \boldsymbol{b}) = \inf\{\mathscr{D}_{\boldsymbol{S}}(\boldsymbol{c}, \boldsymbol{b}): \boldsymbol{c}(X_n) = c_n\}$ gives us a picture of $\boldsymbol{c}^*$'s epistemic status *from a perspective that we should repudiate on the basis of our evidence*!
- Moral: The fact that a credence function some "minimum divergence from the prior" requirement is, by itself, no reason to prefer that credence function.

TWO PICTURES OF UPDATING

I suspect Pettigrew and Leitgeb were misled to thinking that updating procedures can be justified by minimizing $\mathcal{D}_S(\boldsymbol{c}, \boldsymbol{b})$ by a picture of that treats belief change as a matter of *minimally altering* an existing view to accommodate new evidence (so that the posterior is beholden to the prior for its epistemic status).



A better picture has been suggested by Marc Lange (1999).



Here updating is not dynamic at all. Acquisition of evidence is dynamic, but updating is a kind of epiphenomenon that occurs when an agent, at each time, adopts the credences that are best supported by her evidence *at that time*.

WHEN ARE JEFFREY OR PL-UPDATING JUSTIFIED?

It's entirely a matter of the nature of the prior evidence, the import of the new experience, and how the two combine. The key thing for JC, as Jeffrey understood, is whether the experience forces one to alter one's views about the evidential relationships *within each cell of the partition* affected by the learning experience.

✪ If, for each X_n in the partition, and each Y and Z that entail X_n, the new evidence does not alter the probability of Y conditional on Z, then JC is appropriate.

- After a Jeffrey experience, the agent comes to knows that her prior unconditional credences are no longer the best exemplification of the virtues **S** encodes. Indeed, she knows that the best exemplification is some probability for which $\boldsymbol{c}(X_n) = c_n$!

- But, the new evidence will not show that the prior was wrong about everything. In particular, it might not convey any new information relevant to the probabilities of events *conditional on various* X_n.

- In such a case, ✪ will hold because the evidence for setting credence ratios *within each cell of the partition* will be the same before and after the experience.

There are many such cases: In the Potter/Weasley example, suppose Dumbledore tosses a coin and sends Harry to meet you if heads and sent an owl to the Weasleys if tails. In the latter event, Mrs. Weasley who has already tossed two coins, sends Ron if they came up different, Fred if they both came up heads, and George if they came up tails. Since (barring strange scenarios) you'll regard the outcome of Mrs. Weasley's toss as causally independent of the color of the approaching person's hair, accommodating the new data about the red hair will not undermine any of your conditional probabilities like "if it's Ron or Fred then it's twice as likely to be Ron". Thus, aligning your credences to your evidence both before and after the experience will look just like Jeffrey conditioning.

I've not been able to imagine a case in which PL-updating seems appropriate, but that might be just a lack of imagination.

References

Gibbard, Allan [2008] “Rational Credence and the Value of Truth” in Tamar Gendler and John Hawthorne, Eds, *Oxford Studies in Epistemology*, Volume 2. Oxford: Oxford University Press.

Hajek, Alan [2008] “Arguments for–or against–Probabilism?,” *B JPS* **59**: 793-819.

Joyce, James M.. [1998] “A Non-Pragmatic Vindication of Probabilism,” *Philosophy of Science* **65**: 575-603.

Joyce, James M. [2009] “Prospects for an Alethic Epistemology of Partial Belief,” in F. Huber and C. Schmidt-Petri, eds., *Degrees of Belief*: Springer, *Synthese* Library, v.342

Lange, Marc [1999]. "Calibration and the epistemological role of bayesian conditionalization," *Journal of Philosophy* **96** (6):294-324.

Leitgeb, H. and Pettigrew, R. [2010] “An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy,” *Philosophy of Science* 77: 236-272

Lieb, E. H., Osherson, D., Predd, J., Poor, V., Kulkarni, S., Seiringer, R. “Probabilistic Coherence and Proper Scoring Rules.” arXiv:0710.3183, IEEE T. Inform. Theory (in press)

Lindley, David [1982] “Scoring Rules and the Inevitability of Probability,” *International Stat Review* **50**: 1-26.

Maher, Patrick. [2002] “Joyce’s Argument for Probabilism,” *Philosophy of Science* **96**: 73-81.

Savage, L. J. (1971) “Elicitation of Personal Probabilities,” *Journal of the American Statistical Association* **66**: 783-801.

Schervish, Mark J. [1989] “A General Method for Comparing Probability Assessors,” *The Annals of Statistics* **17**: 1856-1879.

Schervish, Mark J., Seidenfeld, Teddy, Kadane, Joseph B. [2009] “Proper Scoring Rules, Dominated Forecasts, and Coherence,” Decision Analysis, Published online in *Articles in Advance*, October 13, 2009 http://da.journal.informs.org/cgi/content/abstract/deca.1090.0153v

Greaves, H. and Wallace, D. [2006] “Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility,” *Mind 105:* 607-632.