Brit. J. Phil. Sci. **63** (2012), 851–857

Calibration and Convexity: Response to Gregory Wheeler

Jon Williamson

ABSTRACT

This note responds to some criticisms of my recent book *In Defence of Objective Bayesianism* that were provided by Gregory Wheeler in his 'Objective Bayesian Calibration and the Problem of Non-convex Evidence'.

In his 'Objective Bayesian Calibration and the Problem of Non-convex Evidence', Gregory Wheeler criticizes the principle I invoke in (Williamson [2010]) for calibrating degrees of belief with chances. Bayesian epistemologists commonly appeal to some sort of calibration norm that says that degrees of belief should be calibrated to known chances. Typically, they invoke a principle variously known as the Straight Rule, Miller's Principle, or the Principal Principle, which says that if one's evidence $\mathscr{E}$ contains the claim that the chance of proposition θ is x, then one should set one's degree of belief in θ relative to that evidence, $P_{\mathscr{E}}(\theta)$, to x, as long as one doesn't have further evidence that trumps or contradicts the chance statement.[1] This type of principle calls for generalization because often we don't have precise chance values as evidence, but constraints on chances—for example, the constraint that the chance of θ is at least 0.7. In (Williamson [2010]), I offer the following generalization:

$$C: P_{\mathscr{E}} \in \langle \mathbb{P}^* \rangle \cap \mathbb{S}.$$

Here $\mathbb{P}^*$ is the set of chance functions that are compatible with evidence $\mathscr{E}$; $\langle \cdot \rangle$ is the convex hull operator (i.e. if $P, Q \in \langle X \rangle$ then $\lambda P + (1-\lambda)Q \in \langle X \rangle$ for each $\lambda \in [0,1]$); and $\mathbb{S}$ is a set of probability functions that satisfy *structural* constraints imposed by evidence $\mathscr{E}$—constraints on degrees of belief that aren't mediated by chances (I argue that structural constraints are generated by evidence of qualitative causal connections or hierarchical relationships, for instance.) In (Williamson [2010], Section 3.3), it is argued that a Calibration

[1] Here, for the sake of brevity, we shall restrict our attention to single-case chance. See (Williamson [2010], Sections 3.3, 10.2), and (Williamson [2011b]) on calibrating degrees of belief to generic frequencies or propensities.

doi:10.1093/bjps/axs003

Advance Access published on May 3, 2012

norm such as the Straight Rule can be motivated by the desire to minimize long-run loss, or the desire to minimize worst-case expected loss, when betting according to one's degrees of belief. This version of the calibration norm invokes the convex hull operator, because one avoids sure loss in the long run just when one bets according to degrees of belief that fall inside the convex hull of the set of chance functions that are compatible with evidence (see below). Convexity becomes intuitively plausible when one considers that one knows, for any proposition, θ, about the past (e.g. the proposition that Aristotle was born on a Thursday), that its chance is now 0 or 1: it is implausible to suggest that one should believe θ to degree 0 or 1 in the absence of further evidence; at the very least, some non-extreme degree of belief such as 1/2 or 1/7 should be deemed rational; such values are in $\langle \mathbb{P}^* \rangle$ but not in $\mathbb{P}^* = \{0, 1\}$.

In Section 2 of his paper, Gregory Wheeler objects that:

> Since there are different ways to parameterize a set of chance functions, and these different parameterizations yield different closed convex hulls, it is misleading to refer to *the* convex hull of $\mathbb{P}^*$; yet, OBE [i.e. objective Bayesian epistemology] is silent on how to construct the correct one.

In fact, though, far from falling silent on this question, my book advocates one particular parameterization throughout, and says this:

> One can define a convex hull of a set of probability functions in various ways, depending on how the probability functions are themselves parameterised. Given the way the Probability norm was introduced [...] it is natural (and quite standard) to define $R = \lambda P + (1 - \lambda)Q$ by $R(\omega) = \lambda P(\omega) + (1 - \lambda)Q(\omega)$. But one might define a probability function in other ways than by its values on the atomic states ω of the form $\pm A_1 \wedge \ldots \wedge \pm A_n$. For example one might define a probability function P by the values it gives to $P(\pm A_i | \pm A_1 \wedge \ldots \wedge \pm A_{i-1})$ for $i = 1, \ldots n$. In general one can not expect different parameterisations to yield the same convex hulls (see, e.g., Haenni et al., 2011, §8.2.1). One of the advantages of the standard parameterisation (which appeals to atomic states) is that it does yield C1 [which says that if $\theta \in \mathcal{E}$ and $\mathcal{E}$ is consistent then $P_{\mathcal{E}}(\theta) = 1$]: if P and Q give probability 1 to θ then so will any convex combination of P and Q. (Williamson [2010], p. 45)

Here $A_1, \ldots, A_n$ are the atomic propositions of the propositional language in question. The point is that different parameterizations are fit for different purposes. The parameterization proposed by Wheeler in terms of $P(\pm A_i | \pm A_1 \wedge \ldots \wedge \pm A_{i-1})$ for $i = 1, \ldots n$ is especially suited to characterizing probability functions in the context of probabilistic networks. This is because probabilistic networks are used to chart probabilistic independencies, and this parameterization both (i) makes probabilistic independencies perspicuous and (ii) can be used to ensure that a convex combination of two probability functions that both satisfy a given independency will also satisfy that independency (see Haenni et al. [2011], section 8.1.2). On the other hand, I would argue that the more typical parameterization in terms of $P(\pm A_1 \wedge \ldots \wedge \pm A_n)$, which is advocated in (Williamson [2010]), is better suited to the context of Bayesian epistemology. This is because this parameterization both (i) fits the motivation in terms of betting alluded to above, and (ii) can be used to ensure that a convex combination of two probability functions that both satisfy a constraint of the form $P(\theta) = x$ also satisfies the same constraint. This is desirable because it is reasonable to insist that if you know that the chance of θ is x then you ought to believe θ to degree x (the Straight Rule).

To see the difference between these two parameterizations, consider the case of three atomic propositions A, B, and C, and evidence which picks out the set of possible chance functions $\mathbb{P}^* = \{P_1^*, P_2^*\}$, where P_1^* gives probability 0.1 to $A \wedge B \wedge C$ and probability 0.9 to $\neg A \wedge \neg B \wedge \neg C$, while P_2^* gives probability 0.1 to $\neg A \wedge B \wedge \neg C$ and probability 0.9 to $A \wedge \neg B \wedge C$. These two possible chance functions share some features: they both give probability 1 to the proposition $A \leftrightarrow C$ and they both render A and C probabilistically independent conditional on B. If one were to form the convex hull of $\mathbb{P}^*$ in a probabilistic-network coordinate system—using the parameters $P(\pm A)$, $P(\pm B|\pm A)$, $P(\pm C|\pm A \wedge \pm B)$—then every convex combination of P_1^* and P_2^* would render A and C probabilistically independent conditional on B, but no proper convex combination would give probability 1 to $A \leftrightarrow C$. On the other hand, if one were to form the convex hull $\langle \mathbb{P}^* \rangle$ in the atomic-state coordinate system—using the parameters $P(\pm A \wedge \pm B \wedge \pm C)$—then every convex combination would give probability 1 to $A \leftrightarrow C$ but no proper convex combination would render A and C probabilistically independent conditional on B. Contra Wheeler, in my book I am not silent as to which parameterization to prefer, since I explicitly adopt the latter, atomic-state parameterization throughout the book. Chance constraints of the form $P^*(A \leftrightarrow C) = 1$ should, I argue, carry over to degrees of belief (via straightforward applications of the Straight Rule), whereas independence constraints of the form $P^*(\pm C|\pm A \wedge \pm B) = P^*(\pm C|\pm B)$ need not. Indeed, in Section 3, Wheeler acknowledges that independence constraints should not carry over to degrees of belief: tosses of a coin that is known to be biased may be probabilistically independent with respect to chance but, of course, the outcome of the first toss conveys useful information about the outcome of the second toss, so the tosses need not be probabilistically independent with respect to rational degree of belief.

*

We can then see that Wheeler's first objection lacks bite. Wheeler claims that Williamson ([2010]) falls silent on the correct way of parameterizing probability functions when taking convex hulls, when in fact the book explicitly adopts one particular parameterization throughout and argues that parameterization should be preferred to the parameterization that Wheeler suggests but later (rightly) rejects. So, while we disagree about whether the book falls silent on this question, it appears that we agree that the $P(\pm A_i | \pm A_1 \wedge \ldots \wedge \pm A_{i-1})$ parameterization should be avoided.

Later on in section 3, Wheeler casts doubt on a parameterization—such as the atomic-state parameterization advocated in Williamson ([2010])—that does not *insist* that tosses of a biased coin be *dependent* with respect to rational degree of belief. The objection here is that even if the calibration norm leaves open that $P_{\mathscr{E}}(H_2|H_1) \neq P_{\mathscr{E}}(H_2)$, where H_i signifies a head at the *i*-th toss, the objective Bayesian Equivocation norm, which says that degrees of belief should equivocate sufficiently between atomic states, will prevent learning from experience. It is true that equivocation will often yield an independency of the form $P_{\mathscr{E}}(H_2|H_1) = P_{\mathscr{E}}(H_2)$, where such an independency is compatible with the evidence after calibration (see, for example, Williamson [2010], Theorem 6.3). It is also true that, if one were to learn solely by conditionalisation, such an independency would render learning from experience impossible, since learning H_1 would not raise the probability of H_2. But, as stressed in the book, objective Bayesian updating only agrees with conditionalization in certain circumstances. In this case, the objective Bayesian would say that if one learns H_1, one's evidence changes to $\mathscr{E}' = \mathscr{E} \cup \{H_1\}$ and one must then calibrate to this new evidence. Since $\mathscr{E}$ says that the bias is either 0.99 or 0.01 in favour of heads, H_2 has a much higher chance now that H_1 is known and the calibration norm forces $P_{\mathscr{E}'}(H_2)$ to calibrate to this chance information. Hence, learning from experience is possible after all. The interested reader may wish to turn to Williamson ([2011a]) for a detailed discussion of where objective Bayesianism stands with respect to learning from experience and to Williamson ([2011b]) for an extended example of the use of confidence interval estimation methods, in conjunction with the calibration norm, to determine rational degrees of belief such as $P_{\mathscr{E}'}(H_2)$.

Wheeler erroneously suggests that my argument for the calibration norm in terms of betting presumes that the set of chance functions delimited by evidence is convex. This appears to be because he takes $\mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$ to be a set of chance functions. Clearly it is no such thing. The calibration norm C given above and in the book makes it clear that while $\mathbb{P}^*$ is a set of chance functions, $\langle \mathbb{P}^* \rangle \cap \mathbb{S}$ is a set of rational belief functions. Furthermore, that $\mathbb{E}$ is convex is not a presumption but a consequence of the justification of the calibration norm: as mentioned above, we need to appeal to the convex hull $\langle \mathbb{P}^* \rangle$ to avoid sure loss in betting scenarios; $\mathbb{S}$ turns out to be convex because it is generated by equality constraints; the intersection of two convex sets is convex; hence $\mathbb{E}$ is convex.

Wheeler suggests that the avoidance of sure loss argument does not apply when $\mathbb{P}^*$ is not itself convex. To see that it can apply, consider the example that Wheeler puts forward: a biased die where outcomes are independent and identically distributed (iid) with unknown bias $P_1^*(H_i) = 0.01$ or $P_2^*(H_i) = 0.51$. Now, if the bias were known to be that of P_1^* and bets were placed according to some fixed $P_{\mathscr{E}}(H_i) = q > 0.01$, then the agent would be sure that a stake-maker could choose positive stakes S to ensure that the betting loss $\sum_{i=1}^{n}(q - I_{H_i})S$, where I_{H_i} is the indicator function for H_i, is positive for sufficiently large n—i.e. a certain loss (Williamson [2010], Section 3.3).[2] Similarly, if $q < 0.01$ then negative stakes can be chosen for a sure loss. But if the evidence equivocates between P_1^* and P_2^* and $0.01 < q < 0.51$ then there is no sure loss: if stakes are chosen positive then the long-run loss will only be positive if $P^*(H_i) = P_1^*(H_i)$; on the other hand, if stakes are chosen negative then the loss will only be positive if $P^*(H_i) = P_2^*(H_i)$; but of course there is no information to tell between the two. However, if $q > 0.51$ then stakes can be chosen positive and if $q < 0.01$ then stakes can be chosen negative for sure loss. Hence, contra Wheeler, the agent needs to ensure that her betting quotients, q, lie in the convex hull in order to avoid certain loss.

Once degrees of belief are narrowed down to this convex hull, the question arises as to whether some members of this convex hull are more appropriate than others as belief functions. In (Williamson [2010], Section 3.4) it is argued that those belief functions that equivocate sufficiently between the atomic states are most appropriate because they control worst-case expected loss (this is the equivocation norm, mentioned above). I argue there that if the particular loss function in operation is unknown, as it usually is, then logarithmic loss is appropriate as a default loss function. In which case, a belief function is sufficiently equivocal just when it is sufficiently close to the function that gives each atomic state the same probability, where distance to this function is explicated by Kullback–Leibler divergence. Equivalently, a belief function is sufficiently equivocal just when it has sufficiently high entropy. Wheeler takes issue with this when the evidence is asymmetric, as is the case in the above example. I suppose his intuition is that it is best if one's belief function is equidistant between two possible chance functions, but without any further justification it is hard to prefer that choice of belief function over one that is sufficiently equivocal in the sense outlined above. One might try to justify an 'Equidistance' norm by arguing that one should equivocate over which value in the interval [0.01, 0.51] one should choose as one's degree of belief, and average over these possible choices to select the midpoint. But such a move would be doubly undesirable to Wheeler, since it presumes not only convexity but also some higher-order equivocation norm that advocates

[2] As is usual in this sort of betting scenario, the agent and the stake-maker have access to the same evidence. This evidence is assumed fixed, i.e. outcomes of previous tosses are not revealed before a bet on H_i. If outcomes were revealed, the agent and the stake-maker would quickly learn which of the two possible chance functions is the one in operation.

equivocating over degrees of belief in a continuum—significantly more contentious than the standard equivocation norm.

I don't pretend to have said the last word about calibration in (Williamson [2010]): I see the Straight Rule as a first approximation to the correct norm, and I have merely tried to formulate a second approximation in the shape of C. Certainly, more can be done to flesh out or improve upon C (see, for example, Williamson [2011b]) and debates can still be had about the merits of taking convex hulls.[3] But I'm not convinced that Wheeler provides new reasons to drop convexity. And, while I agree with Wheeler that there is scope for argument about how best to parameterize probability functions for the purposes of calibration, I deny Wheeler's charge that I have failed to specify a suitable parameterization.

Acknowledgement

I am very grateful to Gregory Wheeler and an anonymous referee for helpful comments and to the British Academy for supporting this research.

SECL
Philosophy Department
University of Kent
Canterbury CT2 7NF, UK
j.williamson@kent.ac.uk

References

Haenni, R., Romeijn, J.-W., Wheeler, G. and Williamson, J. [2011]: *Probabilistic Logics and Probabilistic Networks*, Dordrecht: Springer.

Kyburg, H. E. Jr and Pittarelli, M. [1996]: 'Set-based Bayesianism', *IEEE Transactions on Systems, Man, and Cybernetics*, **26**, pp. 324–39.

Williamson, J. [2010]: *In Defence of Objective Bayesianism*, Oxford: Oxford University Press.

[3] The example of Kyburg Jr and Pittarelli ([1996], Section IV.C) merits further discussion, for instance. In their example, an agent bets on pairs of iid coin tosses, knowing that $P^*(H) \in [0.1, 0.5]$. Taking the convex hull of functions defined on pairs of outcomes that satisfy this constraint admits probability functions for which outcomes are not independent. And, if an agent were to choose such a function, then she would be prone to a positive expected loss. One might argue that this expected loss is undesirable enough to banish convex hulls from the calibration norm.

As it happens, the objective Bayesian procedure would not choose such a function in this case, since it would choose the function P that equivocates between the atomic states, i.e., the P such that $P(HH) = P(HT) = P(TH) = P(TT) = 0.25$, which is compatible with the constraint and which renders outcomes independent. Moreover, a P that satisfies independence would be chosen whatever the endpoints of the interval in the constraint. So expected loss is not positive in the presence of the Equivocation norm.

Nevertheless, if we set the equivocation norm aside and focus just on calibration, there is clearly scope for further debate here about the merits of convexity.

Williamson, J. [2011a]: ‘An Objective Bayesian Account of Confirmation’, in D. Dieks, W. J. Gonzalez, S. Hartmann, T. Uebel and M. Weber (*eds*), *Explanation, Prediction, and Confirmation: New Trends and Old Ones Reconsidered*, Dordrecht: Springer, pp. 53–81.

Williamson, J. [2011b]: ‘Why Frequentists and Bayesians Need Each Other’, *Erkenntnis*, doi:10.1007/s10670-011-9317-8, Epub ahead of print, 21 August 2011.