Objective Bayesian calibration and the problem of non-convex evidence

Gregory Wheeler

CENTRIA - AI Center, Universidade Nova de Lisboa

Abstract

Jon Williamson's Objective Bayesian Epistemology relies upon a calibration norm to constrain credal probability by both quantitative and qualitative evidence. One role of the calibration norm is to ensure that evidence works to constrain a convex set of probability functions. This essay brings into focus a problem for Williamson's theory when qualitative evidence specifies non-convex constraints.

Jon Williamson (2010) provides a spirited defense of a version of Objective Bayesianism which relies upon a calibration norm to constrain credal probability by evidence. According to this norm, an agent's degrees of belief should be constrained by two types of evidence: quantitative evidence, which directly constrains admissible values of chance functions, and qualitative evidence, such as logical or causal constraints on chance variables, which may indirectly constrain chance functions. Once those constraints are in place, the theory maintains that the agent's degrees of belief should be maximally equivocal between the basic outcomes.

I have discussed the calibration norm with sympathy (Wheeler and Williamson 2011), although in that setting our aim was to reconcile the reference class reasoning machinery of Evidential Probability (Kyburg and Teng 2001) with Williamson's Objective Bayesian Epistemology (OBE), so we naturally played down the differences between the two theories. Nevertheless, I have reservations about viewing OBE as a general theory for rational belief. So, in response to Williamson's call to go whole hog, I would like to bring into focus a problem for OBE that arises when qualitative evidence prescribes non-convex constraints on a set of chance functions.

The mechanics of OBE depend on a convex set of probability functions, and one role the calibration norm plays is to enforce this convexity condition. This can be a reasonable approach when the endpoints of an interval designate upper and lower bounds on admissible degrees of belief and the question to answer is what point within this range represents the most cautious position for an agent to take. That is the question that OBE is set up to answer, and it does so by advising the agent to pick the most equivocal point within the convex hull of admissible options

delineated by the calibration norm. In practice this means that OBE shuns both high and low probabilities within a given set when a more equivocal alternative is available. The problem is that not all evidence fits the OBE mold. In particular, non-convex evidence can reverse the order of what is "extreme" and what is "cautious" to believe, throwing the weight of the theory behind precisely the wrong candidates for rational belief.

What is more, the calibration postulates only give a definition of calibrated sets of probability functions rather than an algorithm for how to construct those sets. As it turns out, applying the calibration norm is a case-by-case exercise, one that depends crucially on judgments about how to model quantitative and qualitative evidential constraints. The instrumental role that interpretation plays, absent a general method for calibrating evidence, threatens to render OBE objective in name only.

As a preview of what is to come, we first recount Williamson's postulates for calibration and then demonstrate that the postulates cannot be interpreted as a procedure for constructing calibrated sets of probability functions. Then we demonstrate that the calibration norm cannot be interpreted as a general norm for characterizing evidence by showing that calibration fails to correctly constrain degrees of belief when evidence is non-convex. Finally, in closing, we observe that despite claims to the contrary, OBE fails to clearly distinguish between known evidence and belief.

1 Formal Preliminaries

The strength of an agent's belief is assumed to be representable by a probability function, P, which is defined in standard form with respect to a probability space $(\Omega, \mathscr{F}, P)$ such that $\mathscr{F}$ is a σ-algebra over a set Ω and $P : \mathscr{F} \longrightarrow [0,1]$ is a probability measure defined on the space $(\Omega, \mathscr{F}, P)$ satisfying

(P1) $P(\Omega) = 1$

(P2) $P(\bigcup_{i=1}^{\infty} X_i) = \sum_{i=i}^{\infty} P(X_i)$, when X_i are countable, pairwise disjoint elements of $\mathscr{F}$.

A *probability structure* is a tuple $M = (\Omega, \mathscr{F}, P, V)$, where $(\Omega, \mathscr{F}, P)$ is a probability space and V is an interpretation function associating each element $\omega \in \Omega$ with a truth assignment on the primitive propositions $A, B, \dots$ in an agent's language $\mathscr{L}$ such that $V(\omega)(A) \in \{1, 0\}$ for each $\omega \in \Omega$ and for every proposition in $\mathscr{L}$.

For each primitive proposition in $\mathscr{L}$, we define $M, \omega \models A$ iff $V(\omega)(A) = 1$ and proceed by induction on the structure of propositional formulas. Since P is defined on events in $\mathscr{F}$ rather than propositions, let $[\![A]\!]_M$ denote the set of outcomes within Ω in M where A is true, which will correspond to a subset of $\mathscr{F}$. The following makes explicit the relationship between propositions in $\mathscr{L}$ and events within M, for arbitrary propositional formulas A and B:

2

i. $[\![A \wedge B]\!]_M = [\![A]\!]_M \cap [\![B]\!]_M$,

ii. $[\![A \vee B]\!]_M = [\![A]\!]_M \cup [\![B]\!]_M$,

iii. $[\![\neg A]\!]_M = \overline{[\![A]\!]_M}$.

This relationship licenses us to abuse notation by writing $P(A)$ for $P([\![A]\!]_M)$. From here on we assume that $\mathscr{L}$ and Ω are finite.

OBE rests on three norms.

> **Probability**: The strengths of an agent's beliefs should be representable by a probability function;

> **Calibration**: The agent's degrees of belief should satisfy constraints imposed by his evidence;

> **Equivocation**: The agent's degrees of belief should otherwise be sufficiently equivocal, i.e., the agent's credal probability should be sufficiently close to the equivocator function that assigns atomic states the same probability.

Our concern here is the Calibration norm. The purpose of Calibration is to spell out how background evidence $\mathscr{E}$ should determine the composition of a set $\mathbb{E}$ of probability functions from which the rational agent's credal probability function $P_{\mathscr{E}}$ is selected. Williamson describes this norm as having two components:

> First, if the evidence [$\mathscr{E}$] implies that the physical chance function P^* on $\mathscr{L}$ lies in some set $\mathbb{P}^*$ of probability functions on $\mathscr{L}$, then the agent's belief function $P_{\mathscr{E}}$ should lie in the convex hull $\langle \mathbb{P}^* \rangle$ of $\mathbb{P}^*$, the smallest set containing $\mathbb{P}^*$ that is closed under convex combinations. . . .
> Second, qualitative evidence of, for example, causal, logical, hierarchical, or ontological structure imposes certain structural constraints which force $P_{\mathscr{E}}$ to lie in a set $\mathbb{S}$ of probability functions on $\mathscr{L}$ that satisfy those constraints (Williamson 2010, p. 28).

It turns out that there are different ways to combine qualitative and quantitative evidence which yield different convex sets of chance functions. So, it is misleading to refer to *the* convex hull of $\mathbb{P}^*$. More to the point, insofar as OBE is committed to first taking the convex closure of a set of chance functions, followed by imposing qualitative constraints on that convex set, problems can arise. We return to this point in Section 3.

The following six conditions define the Calibration norm (2010, pp. 39–48, my notation).

(C1) If $A \in S(\mathscr{L})$ is in the agent's evidence base $\mathscr{E}$ and this evidence $\mathscr{E}$ is consistent, then $P_{\mathscr{E}}(A) = 1$.[1]

---

[1] $S(\mathscr{L})$ denotes the set of sentences of $\mathscr{L}$.

(C2) *Miller's Principle*: If $P^*(A) = x$ is the evidence in $\mathscr{E}$ that is most pertinent to A, and $A \in S(\mathscr{L})$, then $P_{\mathscr{E}}(A) = x$.[2]

(C3) $\mathbb{E} \neq \emptyset$.

Condition C3 stipulates that $\mathbb{E}$ is consistent, and the next two conditions install convexity as the consistency maintenance mechanism.

(C4) If $\mathbb{P}^*$ is most pertinent to $\mathscr{L}$, then $\langle \mathbb{P}^* \rangle \subseteq \mathbb{E}$.

(C5) $\mathbb{E} \subseteq \langle \mathbb{P}^* \rangle$.

Here $\mathbb{P}^*$ is a set of chance functions, $\langle \mathbb{P}^* \rangle$ is a convex set of chance functions, and $\mathbb{P}^*$ is most pertinent to $\mathscr{L}$ "if the agent were to grant that $P^* \in \mathbb{P}^*$, then that piece of evidence would be most pertinent to each A on $\mathscr{L}$" (2010, p. 45). Finally, "taking C4 and C5 together, as long as $\mathbb{P}^*$ is most pertinent to $\mathscr{L}$, $\mathbb{E} = \langle \mathbb{P}^* \rangle$" (2010, p. 45).

Postulates C1 to C5 describe the conditions for regulating quantitative evidence, which is the first component to the calibration principle. The second component concerns qualitative evidence, which Williamson treats as a constraint on a convex set of chance functions:

(C6) $\mathbb{E} \subseteq \mathbb{S}$.[3]

Then, Williamson argues, if "we suppose that structural and chance constraints exhaust the ways in which evidence constrains rational degrees of belief, then [these six postulates] motivate the following explication of a Calibration norm" (2010, p. 47):

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

2 A definition without a procedure

While Williamson's postulates may explicate $\mathbb{E}$, they should not be viewed as specifying a procedure for how to construct $\mathbb{E}$. For suppose the postulates were interpreted as a step-by-step procedure for constructing $\mathbb{E}$, a procedure whereby first a set of chance functions is closed under convex combinations, then further constrained by qualitative evidence. It is possible for this procedure to yield a set $\mathbb{E}$ that either fails to remain convex or fails to remain faithful to the evidence. To illustrate this point, consider the following example.

---

[2]'Most pertinent' is defined as follows: "Given A, evidence $\mathscr{E}'$ in $\mathscr{E}$ is the evidence most *pertinent* to A if it is a minimal set of evidence that screens off all other evidence from A: that is, if $\{P_{\mathscr{E}}(A) : P_{\mathscr{E}} \in \mathbb{E}\} = \{P_{\mathscr{E}'}(A) : P_{\mathscr{E}'} \in \mathbb{E}'\}$ and there is no $\mathscr{E}'' \subsetneq \mathscr{E}'$ such that $\{P_{\mathscr{E}}(A) : P_{\mathscr{E}} \in \mathbb{E}\} = \{P_{\mathscr{E}''}(A) : P_{\mathscr{E}''} \in \mathbb{E}''\}$. Here $\mathbb{E}$ (respectively $\mathbb{E}', \mathbb{E}''$) is the set of probability functions compatible with evidence $\mathscr{E}$ (respectively $\mathscr{E}', \mathscr{E}''$)" (2010, p. 39-40).

[3]I omit the details of Williamson's definition of non-parametric, qualitative constraints, since our discussion will not depend on the details of the construction of qualitative constraints but rather on whether the calibration norm correctly represents qualitative evidence in general.

Suppose there is a trick coin that is either biased 0.99 *heads or biased* 0.99 *tails, but you do not know which. How strongly should you believe that a regular toss of this coin will show heads?*

According to OBE, if this is all you know about the experiment, your evidence for the biased coin landing heads, H, is that either $P_1^*(H) = 0.99$ or $P_2^*(H) = 0.01$. So, by the calibration norm, your credal probability function $P_\mathscr{E}$ should lie in the convex hull of your evidence set, $\mathbb{E}$. The question now is how to construct $\mathbb{E}$.

Suppose that $\mathbb{E}$ is constructed by taking the smallest set containing $\mathbb{P}^* = \{P_1^*, P_2^*\}$ *plus* all convex combinations $\lambda P_1^* + (1 - \lambda)P_2^*$, for $\lambda \in [0, 1]$. Within $\mathbb{E}$ is the distribution $P_{1/2}$, defined by $\lambda = 1/2$. It follows that $P_{1/2}(H) = 1/2$ and, according to the equivocation norm, your degree of belief that a single regular toss of this coin will land heads should be $1/2$.

However, to borrow an example of Richard Jeffrey's (1987),[4] notice what happens if we consider two regular tosses of the coin rather than a single toss. Although $P_1^*(H_1) = P_1^*(H_2) = 0.99$ and $P_2^*(H_1) = P_2^*(H_2) = 0.01$, the measure $P_{1/2}$ fails to preserve independence for the event of seeing both tosses land heads:

$$
\begin{aligned}
P_{1/2}(H_1, H_2) &= \frac{P_1^*(H_1)P_1^*(H_2)}{2} + \frac{P_2^*(H_1)P_2^*(H_2)}{2} \\
&= 0.4901, \text{ but} \\
&\neq 0.5 = P_{1/2}^*(H_1)P_{1/2}^*(H_2).
\end{aligned}
$$

Since $P_{1/2}^*$ in $\mathbb{E}$ fails to preserve independence, then $\mathbb{E}$ does not accurately represent the ways in which evidence constrains rational degrees of belief in the event of the coin landing heads on two tosses.

The problem of reconciling stochastic independence with convex closure conditions for sets of probabilities is not a new one (Levi 1980, Jeffrey 1987, Kyburg and Pittarelli 1996, Schervish et al. 2003, Haenni et al. 2011, Cozman 2011). The problem is that moving from a single distribution to a convex set of distributions introduces a *plurality* of independence concepts, and this splintering of probabilistic independence has ramifications for rational choice (Levi 1980, Seidenfeld and Wasserman 1993, Kyburg and Pittarelli 1996), statistical inference (Walley 1991), and probabilistic logic (Haenni et al. 2011), mainly because some inferences from independence conditions which are perfectly sound in the context of a single probability distribution are fallacious in the context of a set of distributions.

Even so, there is a way around Jeffrey's example, and Williamson is well aware of this (Haenni et al. 2011, Chapter 8). Call two random variables, X and Y, *completely stochastically independent*[5] if for all $P \in \mathbb{P}$,

$$
P(X \in \{0, 1\} \cap Y \in \{0, 1\}) = P(X \in \{0, 1\})P(Y \in \{0, 1\}),
$$

---

[4] Thanks here to an anonymous referee.
[5] This term is due to Teddy Seidenfeld.

which guarantees that

$$P(X \in \{0,1\} \mid Y \in \{0,1\}) = P(X \in \{0,1\}), \text{ when } P(Y \in \{0,1\}) > 0.$$

Then, to characterize a convex mixture $\mathbb{E}$ from P_1^* and P_2^* which treats the two coin tosses as completely stochastically independent, define $P = \lambda P_1^* + (1-\lambda)P_2^*$ such that

$$
\begin{aligned}
\alpha &= P(H_1) = 1 - P(T_1), \\
\beta_1 &= P(H_2 \mid H_1) = 1 - P(T_2 \mid H_1) \\
\beta_2 &= P(H_2) \mid T_1) = 1 - P(T_2 \mid T_1)
\end{aligned}
$$

and $\beta_1 = \beta_2$ (Haenni et al. 2011, p. 89).[6] The coordinates α, β_1, and β_2 give a parameterization of the convex set of join probability distributions for two independent coin tosses, where α is simply the probability that the first toss will land heads, and β_1 and β_2 are the probability that the second toss will land heads given that the first toss landed heads and the probability that the second toss will land heads given that the first toss landed tails, respectively.

While these equational constraints ensure that $\mathbb{E}$ treats the two tosses as independent, thus allowing us to steer around Jeffrey's example, the price paid is that the OBE postulates for calibration cannot be viewed as a procedure for how to constrain degrees of belief by evidence. Because 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 is silent on how to construct the correct one.

3   Independence and epistemic relevance

The question of how to construct a calibrated set might be a secondary concern if OBE were to provide the correct characterization of how evidence should constrain degrees of belief. After all, the problem raised in the last section is that some closed convex set of chance functions fail to satisfy all the evidential constraints, not that there are two different ways to satisfy all the calibration postulates. However, in this section we argue that OBE cannot be relied on to give the correct specification of calibrated evidence. The problem is that OBE does not correctly handle nonconvex evidential constraints.

To illustrate the problem, consider once again the biased coin example from before. In addition to the pair of tosses being stochastically independent, another feature of the example is that the evidence about the experiment is not convex: the coin is either strongly biased heads or strongly biased tails. This constraint is useful information that the calibration norm nevertheless throws out by design.

Suppose we represent that the tosses are stochastically independent as above, replacing the two element set $\mathbb{P}$ by a convex hull $\langle \mathbb{P} \rangle$ on which the pair of tosses

---

[6] This example is due to Jan-Willem Romeijn.

are completely stochastically independent. Then, this parameterization effectively replaces an informative constraint on degrees of belief for a second toss, given the outcome of the first, by no constraint at all. Because even though the tosses are stochastically independent, observing the outcome of the first toss is nevertheless epistemically relevant to your estimate of seeing heads on the second toss: given the outcome of the first toss, the probability that the second toss will match is 0.99. But if the convex hull $\langle \mathbb{P} \rangle$ models the tosses as completely stochastically independent, then there is no way for the agent to learn about the bias from observing the outcome of the first toss.

Williamson, discussing a slightly different biased-coin example presented in (Kyburg and Pittarelli 1996), where either $P_1^*(H) = 1$ or $P_2^*(H) = 0$, appears to bite this bullet:

> Clearly, if H has chance 0 or 1, then betting at rate $1/2$ on H over and again may lead to loss (though not sure loss: ... there is no scope for a Dutch book argument here). But that is no reason to preclude initially believing H to degree $1/2$ (Williamson 2010, p. 45, my notation).

We will return to this avoidance of sure loss argument in a moment. For now, notice that even if the objective Bayesian manages to avoid certain loss in the long-run, he does so by stubbornly suffering a loss every step of the way.

Worse still, suppose we set out to moderate the objective Bayesian's stubbornness by allowing him to learn from tosses of the biased coin. A first step would be to view the events as dependent after all. Never mind that this would run counter to the view that $\mathbb{E}$ is a set of chance functions, since we are now considering the structure of the agent's credal state rather than the evidence from the experiment. Set aside that the calibration postulates are not sufficient to characterize a unique $\mathbb{E}$, since they pertain to chance functions rather than epistemic states. Ignore for a moment what principles OBE would use for learning in place of conditionalization. Notice that this learning procedure would still have to reckon with a non-zero probability that the second outcome will not match the first. This means that the calibration norm will still yield a working interval of $[0.01, 0.99]$. Since OBE scores the two original probabilities, $P_1^*(H_2 \mid H_1) = 0.99$ and $P_s^*(H_2 \mid H_1) = 0.01$, as the most extreme candidates within $\mathbb{E}$, even if OBE were to restore learning by discounting the bias towards one side as evidence mounted from observing outcomes of the other, the equivocation norm would nevertheless pull against converging to either $P_{\mathscr{E}}(H) = 0.99$ or $P_{\mathscr{E}}(H) = 0.01$ for any finite sequence of tosses. This misalignment only disappears asymptotically, which means that OBE will have it exactly backwards every step of the way, mistaking the most reasonable answer for the least reasonable option available.

Turn now to Williamson's avoidance of sure loss argument. Williamson concedes that equivocal degrees of belief may open an agent to loss, but he argues that this OBE policy is justifiable on the grounds of minimizing worst-case expected

loss on average (Williamson 2010, p. 66ff.). However, his argument for minimizing average expected loss is made assuming a sample of convex evidence sets, by hypothesis, where the extremal points of each convex set of chance functions are interpreted as defining a polytope *within which* the true chance function lies. But this is precisely the assumption which non-convex evidence rules out. Thus, the avoidance of sure loss argument does not apply. So, by ignoring non-convex evidential constraints and replacing them by closed convex sets, OBE gets it right in at most two cases: initially, when the agent has no information about which way the coin is biased, and here only if the biases are symmetric,[7] and asymptotically, when there is no other option left standing. For all points in between these two limits, which is precisely where we should expect the theory to explain how evidence constraints rational belief, Objective Bayesian Epistemology will recommend the least prudential belief possible.

4   Conclusion

The normative force of an OBE inference hinges on evidence correctly constraining the agent's admissible degrees of belief, since this will define the options over which the agent is expected to equivocate. Nevertheless, applying the calibration norm is a subtle affair, with judgment playing an instrumental role in deciding how to model quantitative and qualitative evidential constraints. Yet, even with this leeway, the theory cannot correctly model non-convex evidence.

In closing, there is another point that falls out from this discussion. Despite claims to the contrary (Williamson 2007, pp. 176–7), OBE does not clearly separate the task of modeling an agent's rational epistemic state from modeling the evidence which justifies that state. As we observed, the non-convex evidence about the biased coin experiment is informative, allowing an agent to learn from previous tosses to quickly make an accurate estimate of the probability of seeing heads on subsequent tosses. But, OBE cannot exploit this information in its modeling of the evidence, and it cannot do so because the structure of the evidence clashes with the machinery for representing the agent's epistemic state.

---

[7]Consider: an initial credence of $1/2$ would not be reasonable if we were told instead that the coin was either biased heads 0.01 or 0.51.

References

Cozman, F. (2011). Sets of probability distributions, independence, and convexity. Forthcoming in *Synthese*.

Haenni, R., Romeyn, J.-W., Wheeler, G., and Williamson, J. (2011). *Probabilistic Logic and Probabilistic Networks*. Synthese Library. Springer, Dordrecht.

Jeffrey, R. (1987). Indefinite probability judgment: A reply to Levi. *Philosophy of Science*, 54:586–591.

Kyburg, Jr., H. E. and Pittarelli, M. (1996). Set-based Bayesianism. *IEEE Transactions on Systems, Man and Cybernetics A*, 26(3):324–339.

Kyburg, Jr., H. E. and Teng, C. M. (2001). *Uncertain Inference*. Cambridge University Press, Cambridge.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, Cambridge, MA.

Schervish, M. J., Seidenfeld, T., Kadane, J. B., and Levi, I. (2003). Extensions of expected utility theory and some limitations of pairwise comparisons. In *Proceedings of the Third Intenational Symposium on Imprecise Probabilities: Theories and Applications*, pages 496–510. Carleton Scientific.

Seidenfeld, T. and Wasserman, L. (1993). Dilation for sets of probabilities. *The Annals of Statistics*, 21:1139–154.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Wheeler, G. and Williamson, J. (2011). Evidential probability and objective Bayesian epistemology. In Bandyopadhyay, P. and Forster, M., editors, *Handbook of the Philosophy of Statistics*. Elsevier Science.

Williamson, J. (2007). Motivating objective Bayesianism: From empirical constraints to objective probabilities. In Harper, W. and Wheeler, G., editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* College Publications, London.

Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford University Press.