

# Invariant Equivocation

Anonymised for submission to Erkenntnis

Draft of July 14, 2016.

## Abstract

Objective Bayesians hold that degrees of belief ought to be chosen in the set of probability functions calibrated with one's evidence. The particular choice of degrees of belief is via some objective, i.e., not agent-dependent, inference process that, in general, selects the most equivocal probabilities from among those compatible with one's evidence. Maximising entropy is what drives these inference processes in recent works by Williamson and Masterton though they disagree as to what should have its entropy maximised. With regard to the probability function one should adopt as one's belief function, Williamson advocates selecting the probability function with greatest entropy compatible with one's evidence while Masterton advocates selecting the expected probability function relative to the density function with greatest entropy compatible with one's evidence. In this paper we discuss the significant relative strengths of these two positions. In particular, Masterton's original proposal is further developed and investigated to reveal its significant properties; including its equivalence to the centre of mass inference process and its ability to accommodate higher order evidence.

Keywords: Maximum Entropy, Objective Bayesianism, Inference Process, Centre of Mass, Language Invariance

## 1 Introduction

“How should one form graded beliefs?” is a question that has long fascinated philosophers. The answer to this question is highly relevant throughout science, law, operational research and policy-making. Intuitively, it is obvious that one's evidence ought to matter when forming beliefs, whether graded or binary. The best way of caching out this ubiquitous intuition is, however, a matter of lively disagreement. The main philosophical protagonists in this debate are: the subjective Bayesians, who can be further subdivided into the radical (de Finetti [5]) and empirical (Lewis [15]) varieties; the objective Bayesians (Jaynes [10], Williamson [27]) and those who favour imprecise probabilities (Joyce [11], Gaifman [6]).

This paper is concerned with objective Bayesianism. Objective Bayesians, like their subjective brethren, hold that one's degrees of belief ought to be consistent with one's evidence. However, in cases where the evidential constraints are satisfied by more than one probability function the objective Bayesians differ from the subjective Bayesians in holding that the choice of a particular probability function from those consistent with the evidential constraints ought to made

in an objective—i.e., agent-independent—manner. Typically, such a choice is made by applying an inference process that picks from among the probability functions consistent with the evidential constraints that function which is, in some sense, maximally equivocal.<sup>1</sup>

Objective Bayesianism is (still!) a minority view with a relatively small but dedicated group of advocates. Even in this relatively small group there are disagreements. One such recent disagreement is that between Williamson [27] and Masterton [16] on how to understand maximal equivocation.

We shall herein first rehearse Williamson’s account in Section 2.1 and 2.2. Then we improve on Masterton’s approach by reformulating it, Section 2.3. This reformulation reveals Masterton to be advocating a centre of mass inference process. This lays bare a previously unknown and unsuspected connection between the centre of mass inference process and the maximum entropy principle, Section 2.5. This connection not only demonstrates the surprising connection between these two apparently different approaches, we can also use it to efficiently calculate the probability function Masterton advocates adopting, see Section 2.6. Furthermore, we overcome the problems of disjunctive evidence and open evidence which beset Masterton’s original proposal. In Section 3 we show how Masterton’s approach naturally generalises to cases in which an agent possesses higher order evidence.

In Section 4 we study an expansion of the language to include sentence which enable the agent to express such higher order evidence. If no such higher order evidence is available, then Masterton’s approach applied to this more expressive language gives the same degrees of belief on the original language as Masterton’s approach applied to the original language. Surprisingly, the corresponding invariance does not hold for Williamson’s approach. We discuss this and further points of interest in the final part of this paper (Section 5).

## 2 Objective Bayesian Accounts

### 2.1 The Framework à la Williamson

#### 2.1.1 Language

Throughout, we consider an agent with a fixed language  $\mathcal{L}$  generated from a finite set of  $n$  propositional variables  $\{A_1, A_2, \dots, A_n\}$ , the standard logical connectives  $\neg, \wedge, \vee, \rightarrow$  and  $\leftrightarrow$ .<sup>2</sup> Let  $\mathcal{S}\mathcal{L}$  be the set of all sentences that can be built from  $\mathcal{L}$  using the logical connectives. A state (or elementary event)  $\omega$  is a sentence of the form  $\pm A_1 \wedge \pm A_2 \wedge \dots \wedge \pm A_n$ , where  $+A_i := A_i$  and  $-A_i := \neg A_i$ . The set  $\Omega$  of such states (elementary events) has cardinality  $N := 2^n$ . The set of the subsets of  $\Omega$ , denoted by  $\mathcal{P}\Omega$ , is the set of propositions (or events).

<sup>1</sup>In awkward cases “maximally equivocal” is replaced by “sufficiently equivocal”, but for the most part we shall ignore this subtlety.

<sup>2</sup>To avoid unnecessary complications we here work over a propositional language; rather than a language of first-order logic.

### 2.1.2 Probabilities

The set of probability functions  $\mathbb{P}$  is identified with

$$\mathbb{P} = \{P : \mathcal{P}\Omega \longrightarrow [0, 1] : \sum_{\omega \in \Omega} P(\omega) = 1 \\ \& P(X) + P(Y) = P(X \cup Y) \ \forall X, Y \in \mathcal{P}\Omega \text{ with } X \cap Y = \emptyset\} . \quad (1)$$

Notation is abused in the usual way, writing  $P(\omega)$  as shorthand for  $P(\{\omega\})$ . The probability of an arbitrary sentence  $\varphi \in \mathcal{S}\mathcal{L}$  is then defined as  $P(\varphi) := \sum_{\substack{\omega \in \Omega \\ \omega \models \varphi}} P(\omega)$ .

The set of probability functions  $\mathbb{P}$  is characterised by the axioms of probability which require the following for all functions  $P : \mathcal{S}\mathcal{L} \longrightarrow [0, 1]$  and all sentences  $\chi, \varphi, \theta \in \mathcal{S}\mathcal{L}$ :

**P1:** If  $\models \chi$ , then  $P(\chi) = 1$ .

**P2:** If  $\models \neg(\theta \wedge \varphi)$ , then  $P(\varphi) + P(\theta) = P(\varphi \vee \theta)$ .

### 2.1.3 Evidence

An agent’s evidence  $\mathcal{E}$  in some context is whatever they rationally take for granted in that context; it need not be expressible in the agent’s language, nor even be part of what the agent knows. We assume that evidence comes in two types<sup>3</sup>: all evidence that “imposes quantitative equality constraints on rational degrees of belief that are not mediated by evidence of chances” [27, p. 47] is *qualitative* evidence, all such evidence that is mediated by evidence of chances is quantitative. For instance, the evidential independence of  $\theta$  and  $\psi$ — $P(\theta|\psi) = P(\theta)$ —is a piece of qualitative evidence, whereas the stochastic independence of these sentences— $ch(\theta|\psi) = ch(\theta)$ —is a piece of quantitative evidence.

At [27, pp. 42-43] Williamson gives an account of how to translate evidence  $\mathcal{E}$  into constraints on rational belief by computing a set  $\mathbb{P}^* \subseteq \mathbb{P}$  consisting of the set of probability functions that are *calibrated* with the evidence  $\mathcal{E}$ . Roughly, this account has  $\mathbb{P}^*$  as the set of probability functions left epistemically open when one’s evidence is exhausted by  $\mathcal{E}$ .

## 2.2 Williamson’s Norms

Williamson put forward three norms that jointly govern the choice of the agent’s degrees of belief. These norms have appeared in a number of works and have remained virtually stable. We here refer to their latest published version for propositional languages, which we take from [13].

**Probability.** The strengths of an agent’s beliefs should satisfy the axioms of probability. That is, there should be a probability function  $P_{\mathcal{E}} : \mathcal{S}\mathcal{L} \longrightarrow$

<sup>3</sup>Both ourselves and Williamson freely allow that there may be other types of evidence and that these will place their own constraints on reasonable credence. Hence, we allow that the calibration norm as explicated herein may be incomplete. This does not make what is presented here unsound, it merely means that it may not be the full story on calibration.

$[0, 1]$  such that for each sentence  $\theta$  of the agent's language  $\mathcal{L}$ ,  $P_{\mathcal{E}}(\theta)$  measures the degree to which the agent, with evidence  $\mathcal{E}$ , believes sentence  $\theta$ . Formally:

$$P_{\mathcal{E}} \in \mathbb{P} .$$

**Calibration.** The strengths of an agent's beliefs should satisfy constraints imposed by her evidence  $\mathcal{E}$ . In particular, if the evidence determines just that physical probability (aka *chance*)  $ch$  is in some set  $\mathbb{P}^*$  of probability functions defined on  $S\mathcal{L}$ , then  $P_{\mathcal{E}}$  should be calibrated to physical probability insofar as it should lie in the convex hull  $\langle \mathbb{P}^* \rangle$  of the set  $\mathbb{P}^*$ . (We assume throughout this paper that chance is probabilistic, i.e., that  $ch$  is a probability function. Furthermore, we restrict attention to non-empty  $\mathbb{P}^*$ .) Formally:

$$P_{\mathcal{E}} \in \langle \mathbb{P}^* \rangle .$$

**Equivocation.** The agent should not adopt beliefs that are more extreme than is demanded by her evidence  $\mathcal{E}$ . That is,  $P_{\mathcal{E}}$  should be a member of  $\langle \mathbb{P}^* \rangle$  that is sufficiently close to the *equivocator* function  $P_{=}$  which gives the same probability to each  $\omega \in \Omega$ , where the *states*  $\omega$  are sentences describing the most fine-grained possibilities expressible in the agent's language.

Much explication and justification of these norms has been offered in the literature. We have nothing to add to this literature here and we feel that both the probability and calibration norms are fully transparent as presented above. We now briefly introduce the equivocation norm in more detail.

### 2.2.1 Equivocation Norm

The *Shannon entropy* of a probability function  $P \in \mathbb{P}$  is defined by

$$H(P) := - \sum_{\omega \in \Omega} P(\omega) \cdot \log(P(\omega)) .$$

A convention we adopt throughout is  $0 \log(0) := 0$ . According to the Maxent version of objective Bayesianism that Williamson adheres to, probability function  $P \in \mathbb{P}$  is *more equivocal* than probability function  $Q \in \mathbb{P}$  if, and only if,  $H(P) > H(Q)$ .

Let  $\Downarrow \langle \mathbb{P}^* \rangle \subseteq \langle \mathbb{P}^* \rangle$  be the set of sufficiently equivocal  $\mathcal{E}$ -calibrated probabilities. Sufficiently equivocal  $\mathcal{E}$ -calibrated probability functions  $P$  are those whose Shannon entropy  $H(P)$  is greatest or, if such do not exist, those whose entropy meets some pragmatically decided threshold. Thus, if

$$\Downarrow \langle \mathbb{P}^* \rangle = \left\{ P_{\mathcal{E}} \in \langle \mathbb{P}^* \rangle : H(P_{\mathcal{E}}) = - \sum_{\omega \in \Omega} P_{\mathcal{E}}(\omega) \log(P_{\mathcal{E}}(\omega)) \text{ is maximized} \right\} \neq \emptyset,$$

then  $\Downarrow \langle \mathbb{P}^* \rangle = \Downarrow \langle \mathbb{P}^* \rangle$ .

If  $\downarrow \langle \mathbb{P}^* \rangle = \emptyset$ , then  $\downarrow \langle \mathbb{P}^* \rangle$  is decided pragmatically by setting a level of entropy that is sufficiently exclusive. Then the equivocation norm in our setting is formally:

$$P_{\mathcal{E}} \in \downarrow \langle \mathbb{P}^* \rangle .$$

This recipe may fail to determine a unique  $P_{\mathcal{E}}$  —  $\downarrow \langle \mathbb{P}^* \rangle$  may not be a singleton — and to the extent that this is so, some room for subjective preference remains [27, p. 158] even where credences are relativised to evidence and language. Arguably, the objective Bayesian can accept this by maintaining that objectivity is not binary, but rather comes in degrees, and that objective Bayesianism is simply a more objective version of Bayesianism than its rivals. This is how Williamson responds to this failure of the equivocation norm to specify a unique credence function in all cases. If  $P_{\mathcal{E}}$  is unique, then we denote this probability function by  $P^\dagger$ .

### 2.3 Masterton

For the time being, we follow [16] by working within the same formal framework as Williamson: i.e., we work on a propositional language and accept the Probability Norm as formulated above. Furthermore, we apply Williamson’s approach to what constitutes evidence and how to turn evidence into constraints on degrees of belief: i.e., how to compute  $\mathbb{P}^*$ .<sup>4</sup> Generally speaking, disagreement arises between these parties with respect to how to understand equivocation, and to a lesser degree, calibration.

The most substantial difference between Masterton and Williamson is over what one should equivocate; Williamson advocates equivocating over the probability functions  $P \in \mathbb{P}^*$  while Masterton advocates equivocating over the probability densities  $\varrho \in \mathbb{C}_{\mathbb{P}^*}^1$  consistent with  $\mathbb{P}^*$ , soon to be described. We shall revisit this disagreement in due course, Section 5, but first we turn to Masterton’s account.

#### 2.3.1 A reformulation of Masterton’s approach

Instead of reiterating Masterton’s original formulation, we give a more efficient and essentially equivalent formulation.

First, we introduce calibrated density distributions  $\varrho$  on the set of probability functions  $\mathbb{P}^*$ . That is, a calibrated density distribution  $\varrho$  is a map from  $\mathbb{P}^*$  to  $\mathbb{R}_{\geq 0}$  such that:

$$\int_{P \in \mathbb{P}^*} \varrho(P) \cdot P \, dP \in \mathbb{P} . \quad (2)$$

The set of such evidence calibrated density distributions  $\varrho$  is denoted by  $\mathbb{C}_{\mathbb{P}^*}^1$ ,

$$\mathbb{C}_{\mathbb{P}^*}^1 := \{ \varrho : \mathbb{P}^* \rightarrow \mathbb{R}_{\geq 0} : \int_{P \in \mathbb{P}^*} \varrho(P) \cdot P \, dP \in \mathbb{P} \} .$$

The meaning of the superscript in  $\mathbb{C}_{\mathbb{P}^*}^1$  shall become clear in Section 4.

<sup>4</sup>Nothing important hinges on this, we are happy with every approach to evidence as long as it results in a set of calibrated functions  $\mathbb{P}^*$ .

The above recipe may fail to be well-defined, if  $\mathbb{P}^*$  is not Lebesgue measurable or if the dimension of  $\mathbb{P}^*$  is strictly less than the dimension of  $\mathbb{P}$ . To address these technical difficulties we will *always* take the Lebesgue measure used for integration to be of the same dimension as  $\mathbb{P}^*$ . For example, if  $\mathbb{P}^*$  is convex and contains more than a single point, then the integral is with respect to the Lebesgue measure of dimension  $\dim(\mathbb{P}^*)$ . For continuum-sized  $\mathbb{P}^*$ , we let  $a$  denote the maximal natural number, such that there exists a set  $U \subseteq \mathbb{P}^*$  of dimension  $a$  which is open in the standard topology of  $\mathbb{R}^a$ . We will also *always* assume that  $\mathbb{P}^*$  is properly Lebesgue measurable in this sense: i.e., the integral  $\int_{P \in \mathbb{P}^*} dP$  is well-defined and strictly between zero and  $+\infty$ . So, if  $\mathbb{P}^*$  is the union of a line segment and a triangle, then  $a = 2$ . We can think of no real life scenarios where the  $\mathbb{P}^*$  fails to be properly Lebesgue measurable, so we deem that this limitation is of negligible practical consequence.

Finally, for finite  $\mathbb{P}^*$  we interpret (2) in this natural sense:

$$\sum_{P \in \mathbb{P}^*} \varrho(P) \cdot P \in \mathbb{P} . \quad (3)$$

That is, we interpret convex combinations of the  $P \in \mathbb{P}^*$  with the density  $\varrho$  specifying the weights for the calibrated probability functions.

The set of density distributions  $\mathbb{C}_{\mathbb{P}^*}^1$  plays a crucial role within Masterton's account. Masterton's equivocation norm requires that one selects the density from this set with greatest entropy. The entropy of a density<sup>5</sup>  $\varrho$  is given by

$$H(\varrho) = - \int_{P \in \mathbb{P}^*} \varrho(P) \log(\varrho(P)) dP .$$

According to Masterton, degrees of belief on  $\mathcal{L}$  have then to be set to

$$P_M^\dagger(\varphi) := \int_{P \in \mathbb{P}^*} \varrho^\dagger(P(\varphi)) P(\varphi) dP , \quad (4)$$

for all sentences  $\varphi \in S\mathcal{L}$  where  $\varrho^\dagger$  is the density in  $\mathbb{C}_{\mathbb{P}^*}^1$  with greatest Shannon entropy. (4) is a well known consequence of the theorem of total probability and Miller's (1966) principle. It implies that the probability of a sentence is the expected probability of that sentence relative to some density function. Masterton's equivocation norm merely asserts that the density function in question should be that calibrated density over  $\mathbb{P}^*$  which is most equivocal. The subscript  $M$  of  $P_M^\dagger$  stands for "Masterton".

In essence, the Lebesgue integral is an ingenious tool to compute weighted combinations of the points of a Lebesgue measurable set. By canonically embedding  $\mathbb{P}^*$  into  $\mathbb{P}$  as a subset, we can thus simply read-off that

$$P_M^\dagger \in \langle \mathbb{P}^* \rangle .$$

Thus Masterton, like Williamson, endorses the calibration norm, but does so for different reasons. Williamson endorses the norm because it is the weakest

<sup>5</sup>Two different densities  $\varrho, \varrho' \in \mathbb{C}_{\mathbb{P}^*}^1$  which only differ on a null-set of  $\mathbb{P}^*$  have the same entropy. However, the expected probabilities with respect to  $\varrho, \varrho'$  are equal: i.e.,  $\int_{P \in \mathbb{P}^*} \varrho(P) \cdot PdP = \int_{P \in \mathbb{P}^*} \varrho'(P) \cdot PdP$ . Null-sets of  $\mathbb{P}^*$  are thus of no interest to us. With some abuse of language we say that the calibrated density with greatest entropy  $\varrho^\dagger$  equals the uniform density on  $\mathbb{P}^*$ . We do this in spite of the fact that there exist densities with the same entropy as  $\varrho^\dagger$ .

constraint that minimises worst case expected logarithmic loss<sup>6</sup> that also offers a non-arbitrary solution to the problem of disjunctive evidence<sup>7</sup>. Masterton endorses the calibration norm because it is a consequence of (4).

### 2.3.2 Alternative densities?

At this point, we want to point out that there is another natural way of setting up densities and equivocating over them. This alternative way is inconsistent with the axiom of probability and hence forbidden in the present Bayesian setting.

One could define densities over the *sentences* of  $\mathcal{L}$ , rather than over the *probability functions* on  $\mathcal{L}$ . One then defines the entropy of a density in the obvious way and computes for all sentences of  $\mathcal{L}$  the density with greatest entropy. The degree of belief in a sentence  $\varphi$  of  $\mathcal{L}$  is then set to the expectation of this density.

For  $\mathbb{P}^* = \mathbb{P}$  and all contingent sentences  $\varphi$  of  $\mathcal{L}$  it holds that every value in  $[0, 1]$  is possible for  $P(\varphi)$ . Thus, the density for  $\varphi$  with greatest entropy is the uniform density over  $[0, 1]$ . It follows that the degree of belief in  $\varphi$  equals  $\frac{1}{2}$ . If  $\Omega$  contains three or more states, then the so-obtained degrees of belief *violate* the axioms of probability,  $\sum_{\omega \in \Omega} P(\omega) = \frac{1}{2} \cdot |\Omega| > 1$ .<sup>8</sup>

## 2.4 Placing Masterton’s and Williamson’s Objective Bayesianism in context

Before moving on to the relationship between Masterton’s Objective Bayesianism and Center of Mass Objective Bayesianism it is useful to reflect on the similarities and differences between Masterton’s Bayesianism and Williamson’s and how they both differ from other variants of Objective Bayesianism. First, the two positions are similar in that they assume the same formal framework, with the sole exception that Masterton includes chance hypotheses in the agent’s language. Second, as stated previously, both Masterton and Williamson are in complete agreement on the probability and calibration norms, though they differ in their reasons for accepting the latter. Third, they both agree that equivocation is about maximising entropy, though they disagree about just what should have its entropy maximised.

An important point with respect to the wider debate is that Masterton’s and Williamson’s endorsement of the calibration norm makes their versions of objective Bayesianism entirely *kinematic*. There are no prior or posterior

<sup>6</sup>Williamson’s [2010, p.64-5] argument here is as follows: If an agent’s probability function  $P$  is not in the convex hull of the  $\mathbb{P}^*$  determined by their evidence, then there is some other probability function  $P' \in \langle \mathbb{P}^* \rangle$  which has a strictly better worst case expected logarithmic loss than  $P$  as shown by [7]; see also [12, 13].

<sup>7</sup>This problem is best exemplified by considering a sentence  $\theta$  that, according to our evidence, is settled one way or the other; so that  $\mathbb{P}^* = \{P : P(\theta) \in \{0, 1\}\}$ . Arguably, restricting credence to any subset of  $\langle \mathbb{P}^* \rangle$  other than  $\mathbb{P}^*$  would be arbitrary, but restricting credence to  $\mathbb{P}^*$  would, by Williamson’s equivocation norm, yield the conclusion that one should either be certain of  $\theta$  or else be certain of its negation. This is highly counterintuitive, as typically one would think that in such a situation one should be as certain in the sentence as its negation, which is the result one obtains by applying the equivocation norm to  $\langle \mathbb{P}^* \rangle$ . Thus no, non-arbitrary, subset of  $\langle \mathbb{P}^* \rangle$  avoids the issues posed by disjunctive evidence, while  $\langle \mathbb{P}^* \rangle$  does avoid those issues.

<sup>8</sup>Degrees of belief in tautologies in  $\mathcal{L}$  are one and degrees of belief in contradictions are zero; which is consistent with the axioms of probability.

credences in their respective Bayesianisms; no new or old evidence. There is simply the evidence the agent has at time  $t$ , which determines the (set of) credence function(s) that it is reasonable/rational for the agent to have at  $t$ . If at time  $t + 1$  the agent's evidence has changed, then this will typically mean that a different (set of) credence function(s) will be reasonable/rational for the agent at  $t + 1$ . No functional relationship between the reasonable credence function of the agent at  $t$  and her credence function at  $t + 1$  is assumed. This makes Masterton's and Williamson's versions of objective Bayesianism substantially different from those dynamic versions, such as Jaynes [9], that seek to identify objective prior probabilities to then conditionalize on new evidence to get objective posterior probabilities. Perhaps most significantly, the problem of old evidence cannot arise in purely kinematic Bayesianism for there is no old or new evidence, just evidence at each moment in time for each agent<sup>9</sup>.

Another agreement between Masterton's and Williamson's objective Bayesianism is that while they agree that the ideal is that an agent's language and evidence should force a unique probability function upon them as reasonable, they both allow that this can fail to happen. The sources of potential failure in this regard are different for the two authors; the source in Williamson's case is that there may be no unique sufficiently equivocal evidence calibrated probability function, while the source in Masterton's case is that there may be no unique sufficiently equivocal evidence calibrated density function. However, both agree that the possibility of such failure in their respective inference processes entails that their Objective Bayesianism is not fully objective. This seems to imply that both Masterton's and Williamson's versions of Objective Bayesianism do not fit well in any of Bandyopadhyay's *et al.* [2] 4 categories of Objective Bayesianism. Neither position is technically 'strongly' objective as they both allow that the inference process may fail to deliver a unique probability (though they are arguably strongly objective in spirit), but nor is either position 'moderately' objective because they do not envisage scientific inference as a problem of deciding between competing theories [3]. Certainly, neither position is a version of a Carnapian [4] style of logical/'necessary' objective Bayesianism. Finally, they are not 'quasi' objective as simplicity plays no role in the inference process and though Masterton and Williamson also allow that the inference process may result in a number of probability functions being equally reasonably they do not condone this fact but rather see it as an unavoidable evil.

---

<sup>9</sup>While Masterton and Williamson avoid the unique probability objection raised by [2], nothing in their previous writings deals with Bandyopadhyay's *et al* criticism that often in science hypotheses are accepted on the strength of the evidence despite their posterior probability in the light of such evidence being very low. While Masterton and Williamson do not have posterior probabilities in their accounts, it is the case that significant evidence for a hypothesis may fail to result in high credence in that hypothesis in their frameworks. Masterton's response to this concern is that while a high degree of belief in a hypothesis is a sufficient condition for its acceptance, it is not a necessary one. Thus, he allows that a hypothesis' acceptance by an agent may be warranted by, e.g., a significant experimental result even when the reasonable degree of belief for that agent on the basis of that evidence is low. That is Masterton, much like van Fraassen [26], holds acceptance and belief to be two entirely distinct doxastic states where warranted acceptance is easier to come by than warranted belief.



## 2.5 Masterton's approach and the centre of mass inference process

We now show how to frame Masterton's approach in terms of the *centre of mass inference process*. Slightly generalising the definition at [18, p. 69] we define an *inference process* to be a map from non-empty sets of probability functions to  $\mathbb{P}$ .

The centre of mass inference process,  $(CM)$ , applied to  $\mathbb{P}^*$ , picks out the probability function  $P_{CM}^\dagger$  at the centre of mass of  $\mathbb{P}^*$ .  $P_{CM}^\dagger$  is usually defined as follows:

$$P_{CM}^\dagger := \frac{\int_{P \in \mathbb{P}^*} P dP}{\int_{P \in \mathbb{P}^*} dP} .$$

Since  $\varrho^\dagger$  is the uniform density satisfying  $\int_{P \in \mathbb{P}^*} \varrho^\dagger(P) dP = 1$  it follows that

$$\begin{aligned} P_M^\dagger &= \int_{P \in \mathbb{P}^*} \varrho^\dagger(P) \cdot P dP \\ &= \frac{\int_{P \in \mathbb{P}^*} P dP}{\int_{P \in \mathbb{P}^*} dP} \\ &= P_{CM}^\dagger . \end{aligned}$$

Thus, Masterton unwittingly advocated adopting the same probability function as the centre of mass inference process picks out.

Unsurprisingly, given the agreement demonstrated above, the basic intuitions cited in support of these inference processes are very similar.

treat all probability functions that satisfies the constraints as equally likely. [...] immediately suggests approximating, or estimating, the true probability by taking the 'average' of all probability function satisfying the constraints. [19, p. 276-277]

Where there is a maximally equivocal element in the evidence-calibrated set of densities over chances, then the reasonable credence to have in  $\theta$ , according to this approach, is the *expected* chance of  $\theta$  relative to [...] that most equivocal density function.[16, p. 422]

## 2.6 Examples

With the centre of mass formulation of Masterton's approach in place we are now in position to compute  $P_M^\dagger$  straight-forwardly using only elementary geometry.

### 2.6.1 $\mathbb{P}^*$ is finite

Example 1: There are only finitely many pairwise different probability functions  $P_1, \dots, P_q$  which are calibrated to the agent's evidence. We have

$$\begin{aligned} \mathbb{P}^* &= \{P_1, \dots, P_q\} \\ \varrho^\dagger(P) &= \begin{cases} \frac{1}{q} & , \text{ if } P \in \{P_1, \dots, P_q\} \\ 0 & , \text{ else } . \end{cases} \end{aligned}$$

$P_M^\dagger$  is then simply the arithmetic mean of the  $P_i$

$$P_M^\dagger = \frac{P_1 + P_2 + \cdots + P_q}{q} .$$

### 2.6.2 Disjunctive Evidence

Example 2: Let  $\mathcal{L}$  contain only a single variable  $A_1$  and let

$$\mathbb{P}^* := \{P \in \mathbb{P} : P(A_1) \in [0.1, 0.3] \cup [0.5, 0.8]\} .$$

The density with greatest entropy  $\varrho^\dagger$  (see Figure 1) thus assigns every  $P \in \mathbb{P}^*$  the value 2. The centre of mass of  $\mathbb{P}^*$ ,  $P_M^\dagger(A_1)$ , then has to satisfy

$$3 \cdot (0.65 - P_M^\dagger(A_1)) = 2 \cdot (P_M^\dagger(A_1) - 0.2) .$$

We find that this constraint is uniquely solved by  $P_M^\dagger(A_1) = 0.47$ .

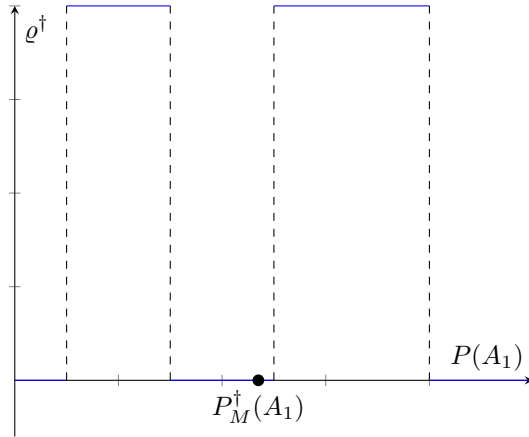


Figure 1: The most equivocal density over  $\mathbb{P}^*$ , with  $P_M^\dagger(A_1)$  marked in the figure.

### 2.6.3 A Trapezium

Example 3: Let  $\mathcal{L} = \{A_1, A_2\}$  and let

$$\mathbb{P}^* = \{P \in \mathbb{P} : P(\omega_1) \in [0.7, 0.9] \& P(\omega_4) = 0\} .$$

$\mathbb{P}^*$  is a trapezium  $T$  with vertices  $P_1 = (0.9, 0.1, 0, 0)$ ,  $P_2 = (0.9, 0, 0.1, 0)$ ,  $P_3 = (0.7, 0.3, 0, 0)$ ,  $P_4 = (0.7, 0, 0.3, 0)$ .

Letting  $P_0 = (1, 0, 0, 0)$  we find that the area of the trapezium  $T$  is the area of the equilateral triangle  $\Delta_3 := \langle P_0, P_3, P_4 \rangle$  minus the area of the equilateral triangle  $\Delta_1 := \langle P_0, P_1, P_2 \rangle$ , see Figure 2.

For the area of  $T$  we now easily find

$$\text{area}(T) = \frac{\sqrt[3]{3}}{4} 0.3^2 - \frac{\sqrt[3]{3}}{4} 0.1^2 = 0.02 \cdot \sqrt[3]{3} .$$

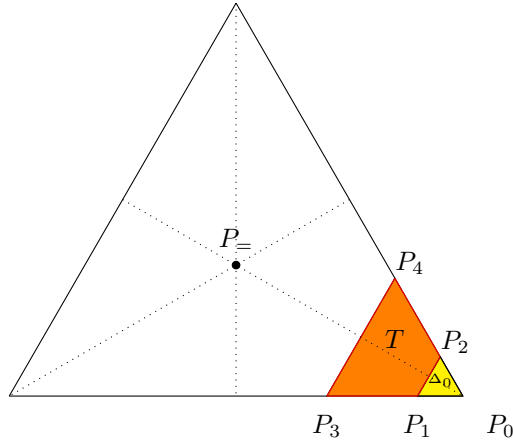


Figure 2: Projection of trapezium onto the plane with  $P(\omega_4) = 0$ . The dotted lines are axis of symmetry.

By the symmetry of  $\mathbb{P}^*$  we can infer that the most equivocal density is invariant under permuting  $\omega_2$  and  $\omega_3$ . We thus infer  $P_M^\dagger(\omega_2) = P_M^\dagger(\omega_3)$  and also that  $P_M^\dagger(\omega_4) = 0$ . We are thus looking for the point  $X = (1-x, \frac{x}{2}, \frac{x}{2}, 0) \in T$  on the line segment  $h$  connecting  $(0.9, 0.05, 0.05, 0)$  and  $(0.7, 0.15, 0.15, 0)$  which is the centre of mass of  $T$ .

Note that  $X$  is not the midpoint of  $h$ . Rather,  $X$  needs to lie on the line segment which connects  $P_x := (1-x, x, 0, 0)$  and  $P^x := (1-x, 0, x, 0)$  which cuts  $T$  into two equally large trapeziums  $T_1$  and  $T_2$

$$\begin{aligned} T_1 &= \langle P_1, P_2, P_x, P^x \rangle \\ T_2 &= \langle P_3, P_4, P_x, P^x \rangle \\ \text{area}(T_1) &= \text{area}(\langle P_x, P^x, P_0 \rangle) - \text{area}(\Delta_0) \\ \text{area}(T_2) &= \text{area}(\Delta_3) - \text{area}(\langle P_x, P^x, P_0 \rangle) . \end{aligned}$$

The geometry is depicted in Figure 3.

We thus obtain the following constraint on  $x$

$$\begin{aligned} \text{area}(T_1) &= x^2 \frac{\sqrt[2]{3}}{4} - 0.1^2 \frac{\sqrt[2]{3}}{4} = 0.3^2 \frac{\sqrt[2]{3}}{4} - x^2 \frac{\sqrt[2]{3}}{4} = \text{area}(T_2) \\ &= \frac{\text{area}(T)}{2} = 0.01 \cdot \sqrt[2]{3} , \end{aligned}$$

which is uniquely solved by  $x = \sqrt[2]{0.05} \approx 22.36\%$ .

We thus find  $P_M^\dagger$  as follows

$$P_M^\dagger = (1 - \sqrt[2]{0.05}, \frac{\sqrt[2]{0.05}}{2}, \frac{\sqrt[2]{0.05}}{2}, 0) \approx (77.64\%, 11.18\%, 11.18\%, 0) .$$

The entropy maximiser  $P^\dagger$  is  $P^\dagger = (0.7, 0.15, 0.15, 0)$ .  $P_M^\dagger$  and  $P^\dagger$  are plotted in Figure 4.



We assume that the agent does not have *any* further evidence; in particular, the agent does not have any further evidence on the chances of heads from coins from a particular machine. What ought the agent's degrees of belief be?<sup>10</sup>

Williamson's account gives  $P^\dagger(\text{Heads}) = 0.5$ , since  $\langle \mathbb{P}^* \rangle = [0.48, 0.72]$  and in Williamson's account one ought to adopt the Shannon entropy maximiser in  $\langle \mathbb{P}^* \rangle$ . Masterton's answer is

$$P_M^\dagger(\text{Heads}) = \frac{4 \cdot 0.5 + 20 \cdot 0.69}{24} = \frac{15.8}{24} = 65.8\bar{3}\% .$$

According to [16] the problem with Williamson's account is that his answer  $P^\dagger(\text{Heads}) = 0.5$  is not influenced by the possibility of the coin being produced by Machine B or Machine C.

A similar problem besets Masterton's own account. Let us modify our example such that there are now fifty machines of type C, one single machine of type A and one single machine of type B. Masterton's, as well as Williamson's, account advocate adopting the same belief function in both examples. This seems peculiar. The coin to be tossed has, with overwhelming likelihood, been produced by a C-type machine. Yet,  $P_M^\dagger(\text{Heads}) = 0.658\bar{3}$  and  $P^\dagger(\text{Heads}) = 0.5$  are incompatible with the possibility that the coin was produced by a C-type machine.

The number of machines of type C *ought* to influence rational degrees of belief, we claim. We now turn to a suggestion for how this might be.

### 3.2 Scenario Equivocation

In this concrete example with 52 coin-producing machines we want to suggest the following approach. First, the agent equivocates over the possibilities regarding which machine produced the coin. Then the agent equivocates over the probability functions in  $\mathbb{P}_A^*$ , respectively  $\mathbb{P}_B^*$  and  $\mathbb{P}_C^*$ . We obtain

$$\begin{aligned} P_{M^+}^\dagger(\text{HEADS}) &= \frac{1}{52} \cdot 50\% + \frac{1}{52} \cdot 63\% + \frac{50}{52} \cdot 76\% \\ &= \frac{50 + 63 + 50 \cdot 76}{5200} = \frac{3913}{5200} \\ &= 0.7525 . \end{aligned}$$

As expected,  $P_{M^+}^\dagger$  is not only calibrated to  $\mathbb{P}_C^*$  but  $P_{M^+}^\dagger$  is close to the centre of  $\mathbb{P}_C^*$ .

In general, we want to suggest the following: if all the agent takes for granted is that she is in one of  $t$  mutually exclusive scenarios, then she ought first to equivocate over this set of possible scenarios. That is, the agent ought to have a degree of belief  $\frac{1}{t}$  for each of those scenarios that it is *her* scenario. Then, in each scenario  $i$  the agent calculates the set of epistemically open probability functions,  $\mathbb{P}_i^*$  for  $1 \leq i \leq t$ . Next, the agent calculates the maximally equivocal

<sup>10</sup>To keep the example simple, we restrict ourselves here to the case in which there is no information as to how likely any given situation is. If such information is available, then it ought to be taken into account, in an appropriate manner. For example, if situation  $i$  is taken to be twice as likely as situation  $k$ , then situation  $i$  should be given double the weight of situation  $k$ . We shall come back to higher order evidence in Section 4. Lewis describes a similar procedure at [15, p. 266].

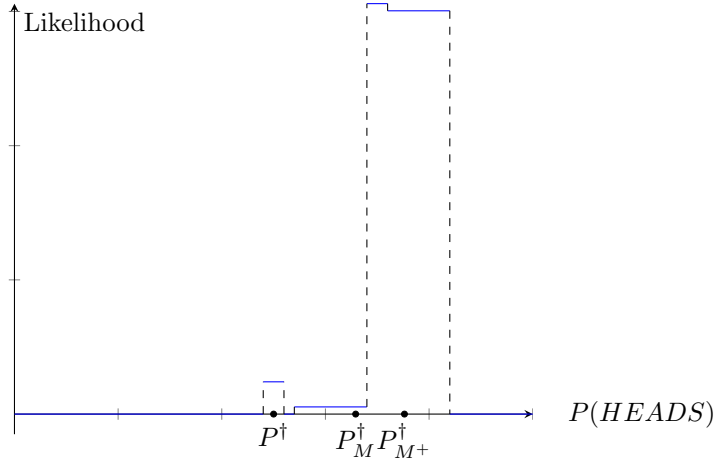


Figure 5: Likelihoods of biases and the value of maximum entropy functions  $P^\dagger$ ,  $P_M^\dagger$  and  $P_{M+}^\dagger$  for *HEADS*.

function in  $\mathbb{P}_i^*$  à la Masterton. Finally, the agent adopts the arithmetic mean of these maximally equivocal functions.

With this picture in mind, we now find

$$\begin{aligned}
 P_{M+}^\dagger &= \frac{1}{t} \cdot \sum_{i=1}^t P_{CM}^\dagger(\mathbb{P}_i^*) \\
 &= \frac{1}{t} \cdot \sum_{i=1}^t \frac{\int_{\mathbb{P}_i^*} P \, dP}{\int_{\mathbb{P}_i^*} dP} \\
 &= \frac{1}{t} \cdot \int_{P \in \mathbb{P}^*} P \cdot \left( \sum_{i=1}^t \rho_i^\dagger(P) \right) dP \\
 &= \int_{P \in \mathbb{P}^*} P \cdot \rho^\dagger(P) \, dP \quad ,
 \end{aligned}$$

where the last formulations are well defined, if and only if all  $\mathbb{P}_i^*$  have the same dimension.  $\rho^\dagger(P)$  for *Heads* in our coin machine example is given in Figure 5. For  $t = 1$  this agrees with Masterton's approach, while generally:

$$P_{M+}^\dagger = \frac{1}{t} \sum_{i=1}^t P_{M_i}^\dagger \quad .$$

An alternative equivocation norm results if one takes the maximally equivocal function in  $\mathbb{P}_i^*$  for each  $i$  à la Williamson:

$$P_+^\dagger := \frac{1}{t} \sum_{i=1}^t P_i^\dagger \quad .$$

In the coin machine example,

$$\frac{1}{t} \sum_{i=1}^t P_i^\dagger = \frac{0.5 + 0.54 + 50 \cdot 0.68}{52} = 0.674 \neq 0.7525 = \frac{1}{t} \sum_{i=1}^t P_{M_i}^\dagger \quad .$$

Indeed, we find that  $P_+^\dagger(\text{Heads})$  is incompatible with the coin being produced by machine  $C$ . Thus, this approach does not overcome the concern we raised for Williamson’s and Masterton’s inference principles.

### 3.3 Centre of Mass Infinity

A possible complication at this point is that centre of mass  $CM$  equivocation over the  $\mathbb{P}_i^*$  may be found objectionable on the grounds that such inference is *not* language invariant. That is, an agent with a larger language  $\mathcal{L}'$ —obtained from  $\mathcal{L}$  by adding further propositional variables—with the same evidence would draw different inferences about the sentences  $\varphi \in \mathcal{S}\mathcal{L}$  than the present agent. General representation dependence of inference is well-known to be virtually unavoidable (see [8]) but language invariance is achievable, and has been achieved by Williamson and those of his ilk.

To the extent that one finds this objection serious one might be tempted to return to equivocating in Williamson’s fashion despite the previously identified concerns. However, there is another response; namely, using centre of mass infinity ( $CM_\infty$ )[23] equivocation over the  $\mathbb{P}_i^*$  instead of straight  $CM$ .<sup>1112</sup> This solves any concerns about language invariance as it has been shown (*ibid*) that  $CM_\infty$  is a language invariant inference process. It follows that if  $CM_\infty$  replaces  $CM$  as the form of equivocation over each  $\mathbb{P}_i^*$ , then the result of the new norm is a convex combination of language invariant processes, which will itself be language invariant. For a longer discussion on these issues see [19, Section 4].

For convex and closed  $\mathbb{P}^*$ , the function picked out by  $CM_\infty$  maximises

$$\sum_{\omega \in \Omega_I} \log(P(\omega)) \tag{5}$$

where  $\Omega_I$  is the subset of states  $\omega$  such that  $P(\omega) > 0$  is consistent with  $\mathbb{P}^*$ , see [18, page 74].<sup>13</sup> The probability function picked out by this inference process is denoted by  $P_{CM_\infty}^\dagger$ . We argue in Section 5 that this option is ultimately unappealing despite it satisfying the desideratum of language invariance.

### 3.4 An Objection of Feasibility

An opponent of the above suggested scenario equivocation might object that it is not always possible for the agent to determine whether her epistemic situation can be neatly split into finitely many distinct scenarios. For example, an agent might take it rationally for granted that  $P^* \in \mathbb{P}_A^* \cup \mathbb{P}_B^* \cup \mathbb{P}_C^*$  but be unsure about why exactly that is. The objection is that we have not spelled out how the agent ought to proceed. Williamson’s account is, of course, immune to this objection.

We concede the point that we have not put forward an approach which covers such cases. We would like to make two points. First, in many applications it will be clear whether there are such scenarios or not. We claim that at least

<sup>11</sup>Another approach satisfying language invariance, via marginalisation of Dirichlet priors, has been taken in [14].

<sup>12</sup>Interest in  $CM_\infty$  has waned in recent years until it very recently resurfaced in [1] and [29].

<sup>13</sup>If  $\omega \in \Omega \setminus \Omega_I$ , then  $P(\omega) = 0$  for all  $P \in \mathbb{P}^*$ . Hence,  $\log(P(\omega)) = -\infty$ . So, would the above sum contain such an  $\omega$ , then the entire sum would have value  $-\infty$ .

in these cases scenario equivocation is appropriate. Second, a case as in the above objection is simply an epistemically “hard” case. We do not have a general approach to “hard” cases at this time. In general, we suspect that the appropriate course of action in such cases is dependent on the exact epistemic situation of the agent; hence a single, general inference process for all such cases is likely a pipe dream.

Before moving on, we consider one related objection which we feel has less merit than the above. An agent faced with the situation we describe in Section 3.1 might split the world into 4 scenarios (rather than three) as follows: in Scenario 1  $P^* \in [0.48, 0.5]$ , Scenario 2  $P^* \in [0.5, 0.52]$ , Scenario 3  $P^* \in [0.54, 0.72]$  and Scenario 4  $P^* \in [0.68, 0.84]$ . This agent would assign  $P^* \in [0.48, 0.52]$  a likelihood of  $0.5 \neq \frac{1}{3}$ . So, our approach gives two different answers. We would argue that in this case the agent’s evidence determines how to split the world into scenarios. Therefore, our approach only gives one answer. We shall revisit such “problems” of language dependence in Section 5.2.

## 4 Chance Invariance

One criticism of CM is the dependence of inferences on the underlying language, i.e., CM is not language invariant. While Williamson’s equivocation procedure is language invariant it suffers from a related problem. It matters whether one equivocates over probability functions or over probability densities. As we saw earlier, equivocation over densities requires one to adopt the centre of mass probability function  $P_{CM}^\dagger$  and not the Shannon entropy maximiser  $P^\dagger$ .

We here show that Masterton’s approach does *not* suffer from a related problem. If one equivocates over densities of some order greater or equal than the order one has evidence for, then one adopts the same probability function for decision making. That is, decisions do not depend on the agent’s language as long as the language is sufficiently rich to allow the formalisation of the agent’s total evidence. To allow for this richer language we now enrich  $\mathcal{L}$  to include densities.

### 4.1 The Base Case

Let us first consider the case in which the agent does *not* have any higher order evidence. That is, the agent has no evidence which favors one of the  $P \in \mathbb{P}^*$  over the others.

We define densities of higher order  $n \geq 1$  as follows:

$$\begin{aligned} \mathbb{C}_{\mathbb{P}^*}^0 &:= \mathbb{P}^* \\ \mathbb{C}_{\mathbb{P}^*}^1 &:= \{f_1 : \mathbb{P}^* \rightarrow \mathbb{R}_{\geq 0} : \int_{P \in \mathbb{P}^*} f_1(P) \cdot P \, dP \in \mathbb{P}\} \\ \mathbb{C}_{\mathbb{P}^*}^2 &:= \{f_2 : \mathbb{C}_{\mathbb{P}^*}^1 \rightarrow \mathbb{R}_{\geq 0} : \int_{f_1 \in \mathbb{C}_{\mathbb{P}^*}^1} f_2(f_1) \cdot f_1 \, df_1 \in \mathbb{C}_{\mathbb{P}^*}^1\} \end{aligned}$$

and so on...

$$\mathbb{C}_{\mathbb{P}^*}^{n+1} := \{f_{n+1} : \mathbb{C}_{\mathbb{P}^*}^n \rightarrow \mathbb{R}_{\geq 0} : \int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} f_{n+1}(f_n) \cdot f_n \, df_n \in \mathbb{C}_{\mathbb{P}^*}^n\} .$$



Intuitively, the  $f_{n+1}$  specify how likely the densities  $f_n$  are. We have already encountered the densities of order 1 in  $\mathbb{C}_{\mathbb{P}^*}^1$  in Section 2.3.1. In Section 2.6, we gave a geometric interpretation of these densities.

We now go on to show that the  $\mathbb{C}_{\mathbb{P}^*}^n$  are Lebesgue measurable. For the remainder of this paper we will always take it that densities are well-enough behaved that the above integrals exist, i.e., are well-defined. We cannot envision a *practical* application in which densities are not Lebesgue measurable.

**Proposition 1.**  $\mathbb{C}_{\mathbb{P}^*}^n$  is convex for all  $n \geq 1$ .

**Proof:** We proceed by induction. For  $n = 1$  and  $f_1, f'_1 \in \mathbb{C}_{\mathbb{P}^*}^1$  we have

$$\int_{P \in \mathbb{P}^*} \frac{f_1 + f'_1}{2}(P) \cdot P \, dP = \frac{1}{2} \left( \int_{P \in \mathbb{P}^*} f_1(P) \cdot P \, dP + \int_{P \in \mathbb{P}^*} f'_1(P) \cdot P \, dP \right) .$$

Since  $\mathbb{P}$  is convex, this function is in  $\mathbb{P}$ .

For general  $n + 1$  and  $f_{n+1}, f'_{n+1} \in \mathbb{C}_{\mathbb{P}^*}^{n+1}$  we have

$$\begin{aligned} & \int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} \frac{f_{n+1} + f'_{n+1}}{2}(f_n) \cdot f_n \, df_n \\ &= \frac{1}{2} \left( \int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} f_{n+1}(f_n) \cdot f_n \, df_n + \int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} f'_{n+1}(f_n) \cdot f_n \, df_n \right) . \end{aligned}$$

By definition of  $\mathbb{C}_{\mathbb{P}^*}^{n+1}$ , both integrals are members of  $\mathbb{C}_{\mathbb{P}^*}^n$ . Since  $\mathbb{C}_{\mathbb{P}^*}^n$  is convex by the induction hypothesis, the function on the right hand side of the equality symbol is a function in  $\mathbb{C}_{\mathbb{P}^*}^n$ . ■

Following the template of Shannon Entropy we define the entropy of higher order densities. The Shannon Entropy of  $P$  is its negative expected logarithmic utility where  $P$  is summed over all states  $\omega \in \Omega$ . Arguably, the entropy of a density  $f^{n+1}$  can be measured by computing its negative expected logarithmic utility. As a  $(n + 1)$ -density applies to  $n$ -densities, so the underlying space is  $\mathbb{C}_{\mathbb{P}^*}^n$ . Since this space is infinite and continuous, the sum is replaced by a uniform integral. Since convex sets are Lebesgue integrable the expressions below are well-defined.

We define the entropy of an  $f_{n+1} \in \mathbb{C}_{\mathbb{P}^*}^{n+1}$  by

$$H(f_{n+1}) := - \int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} f_{n+1}(f_n) \cdot \log(f_{n+1}(f_n)) \, df_n . \quad (6)$$

Though this notion of entropy on a continuous domain might be rather arcane to the formal epistemologist, it is widely used in the applied sciences.

The density  $f_{n+1}^\dagger$  with maximal entropy (modulo null-sets) is constant, i.e., it assigns every density  $f_n \in \mathbb{C}_{\mathbb{P}^*}^n$  the same value. This is in-line with the intuition that maximum entropy functions are maximally equivocal, i.e., assign the same value to all members of the underlying domain.

Computing the expected  $n$ -density with respect to  $f_{n+1}^\dagger$  we obtain

$$\int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} f_{n+1}^\dagger(f_n) \cdot f_n \, df_n . \quad (7)$$

By definition, this is a density  $f_n \in \mathbb{C}_{\mathbb{P}^*}^n$ . Furthermore,  $f_{n+1}^\dagger(f_n)$  is constant on  $\mathbb{C}_{\mathbb{P}^*}^n$ . So, it must be equal to  $\|\mathbb{C}_{\mathbb{P}^*}^n\|^{-1}$ . Hence,

$$\int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} f_{n+1}^\dagger(f_n) \cdot f_n df_n = \frac{1}{\|\mathbb{C}_{\mathbb{P}^*}^n\|} \int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} f_n df_n \quad (8)$$

is the centre of mass of  $\mathbb{C}_{\mathbb{P}^*}^n$ ,  $CM(\mathbb{C}_{\mathbb{P}^*}^n)$ .

On the other hand, we can understand (8) as a function from  $\mathbb{C}_{\mathbb{P}^*}^{n-1}$  to  $\mathbb{R}_{\geq 0}$  which is *constant* on  $\mathbb{C}_{\mathbb{P}^*}^{n-1}$ . But this means that (8) is equal to  $f_n^\dagger$ . We thus obtain

$$CM(\mathbb{C}_{\mathbb{P}^*}^n) = \int_{f_n \in \mathbb{C}_{\mathbb{P}^*}^n} f_{n+1}^\dagger(f_n) \cdot f_n df_n = f_n^\dagger. \quad (9)$$

So, the expected  $n$ -density relative to the most equivocal  $(n+1)$ -density is the most equivocal  $n$ -density. More informally, taking expectations with respect to the maximally equivocal  $(n+1)$ -density yields the centre of mass of  $\mathbb{C}_{\mathbb{P}^*}^n$  which equals the maximally equivocal  $n$ -density.

We already saw a special case of this phenomenon in Section 2.3.1.  $CM(\mathbb{P}^*)$  can be obtained by computing expectation with respect to the maximally equivocal function in  $\mathbb{C}_{\mathbb{P}^*}^1$ , if no first (nor any higher) order evidence is available.

We now find

$$P_{CM}^\dagger = \int_{P \in \mathbb{P}^*} f_1^\dagger(P) \cdot P dP \quad (10)$$

$$= \int_{P \in \mathbb{P}^*} \overbrace{\left( \int_{f_1 \in \mathbb{C}_{\mathbb{P}^*}^1} f_2^\dagger(f_1) f_1 df_1 \right)}^{f_1^\dagger} (P) \cdot P dP \quad (11)$$

$$= \int_{P \in \mathbb{P}^*} \underbrace{\left( \int_{f_1 \in \mathbb{C}_{\mathbb{P}^*}^1} \overbrace{\left( \int_{f_2 \in \mathbb{C}_{\mathbb{P}^*}^2} f_3^\dagger(f_2) f_2 df_2 \right)}^{f_2^\dagger} (f_1) f_1 df_1 \right)}_{f_1^\dagger} (P) \cdot P dP. \quad (12)$$

The upshots are two-fold. On whatever density order one equivocates, as long as one computes expectations as above, one always obtains  $P_{CM}^\dagger$ . A phenomenon we term *chance invariance*.

Furthermore, the centre of mass of  $\mathbb{C}_{\mathbb{P}^*}^n$  and the density with greatest entropy in  $\mathbb{C}_{\mathbb{P}^*}^n$  are one and the same function for  $n \geq 1$ .

## 4.2 Higher Order Evidence

Let us now consider the case in which the agent *does* have higher order evidence. We already considered such a case in the above 52 machine example, see Figure 5. To tackle the general case, we first have to define the sets of evidence

calibrated density functions. We define:

$$\mathbb{D}_{\mathbb{P}^*}^0 := \mathbb{P}^*$$

$$\mathbb{D}_{\mathbb{P}^*}^1 := \{f_1 \in \mathbb{C}_{\mathbb{P}^*}^1 : f_1 \text{ consistent with the agent's evidence}\}$$

$$\mathbb{D}_{\mathbb{P}^*}^2 := \{f_2 \in \mathbb{C}_{\mathbb{P}^*}^2 : f_2(f_1) = 0 \text{ for all } f_1 \notin \mathbb{D}_{\mathbb{P}^*}^1 \text{ \& } f_2 \text{ consistent with the agent's evidence}\}$$

and so on...

$$\mathbb{D}_{\mathbb{P}^*}^{n+1} := \{f_{n+1} \in \mathbb{C}_{\mathbb{P}^*}^n : f_{n+1}(f_n) = 0 \text{ for all } f_n \notin \mathbb{D}_{\mathbb{P}^*}^n \text{ \& } f_{n+1} \text{ consistent with the agent's evidence}\} .$$

Note that  $\mathbb{D}_{\mathbb{P}^*}^n \subseteq \mathbb{C}_{\mathbb{P}^*}^n$ , so we immediately obtain that the entropy of densities in  $\mathbb{D}_{\mathbb{P}^*}^n$  is also well-defined.

For example, in the situation depicted in Figure 5 we have

$$\mathbb{D}_{\mathbb{P}^*}^1 := \{f_1 \in \mathbb{C}_{\mathbb{P}^*}^1 : f_1 = \frac{1}{52}f_A + \frac{1}{52}f_B + \frac{50}{52}f_C$$

$$\text{for some } f_A \in \mathbb{C}_{\mathbb{P}^*_A}^1, f_B \in \mathbb{C}_{\mathbb{P}^*_B}^1, f_C \in \mathbb{C}_{\mathbb{P}^*_C}^1\}$$

$$\mathbb{D}_{\mathbb{P}^*}^2 := \{f_2 \in \mathbb{C}_{\mathbb{P}^*}^2 : f_2(f_1) = 0 \text{ for all } f_1 \notin \mathbb{D}_{\mathbb{P}^*}^1\}$$

and so on...

$$\mathbb{D}_{\mathbb{P}^*}^{n+1} := \{f_{n+1} \in \mathbb{C}_{\mathbb{P}^*}^{n+1} : f_{n+1}(f_n) = 0 \text{ for all } f_n \notin \mathbb{D}_{\mathbb{P}^*}^n\} .$$

We plotted the density in  $\mathbb{D}_{\mathbb{P}^*}^1$  with greatest entropy in Figure 5.

In practical applications, there will be a finite upper bound on the order of the agent's evidence. Let  $N \in \mathbb{N}$  be that bound. Since the agent does not have any higher order evidence, the agent does not have a reason to favor one of the densities in  $\mathbb{D}_{\mathbb{P}^*}^N$  over another such density. Hence, an agent ought to equivocate over these densities.

So, an agent computes the density  $f_{N+1}^\dagger \in \mathbb{D}_{\mathbb{P}^*}^{N+1}$  with maximum entropy. Again, the maximum entropy density assigns all  $f_N \in \mathbb{D}_{\mathbb{P}^*}^N$  the same weight. Hence, the expected  $N$  density with respect to  $f_{N+1}^\dagger$  is the centre of mass of  $\mathbb{D}_{\mathbb{P}^*}^N$ . Armed with this density it is then straight forward to compute the probability used for decision making by computing expectations:

$$\int_{f_1 \in \mathbb{D}_{\mathbb{P}^*}^1} \left( \int_{f_2 \in \mathbb{D}_{\mathbb{P}^*}^2} \dots \left( \int_{f_N \in \mathbb{D}_{\mathbb{P}^*}^N} f_{N+1}^\dagger(f_N) f_N df_N \right) \dots (f_2) f_2 df_2 \right) (f_1) f_1 df_1 \in \mathbb{P}.$$

Now suppose the agent equivocated over some higher order  $N+k$  with  $k > 1$ . The agent would then compute the density in  $\mathbb{D}_{\mathbb{P}^*}^{N+k}$  with maximum entropy,  $f_{N+k}^\dagger$ , being fully indifferent towards these densities. Computing expectations over  $\mathbb{D}_{\mathbb{P}^*}^{N+k-1}$  with respect to  $f_{N+k}^\dagger$  yields the centre of Mass of  $\mathbb{D}_{\mathbb{P}^*}^{N+k-1}$  since  $f_{N+k}^\dagger$  is constant on  $\mathbb{D}_{\mathbb{P}^*}^{N+k-1}$ . Repeating this step  $k-1$ -many further times, the agent obtains  $f_{N+1}^\dagger$ . Thus Masterton's approach generalised to higher order's has the nice property that equivocating on any order greater than that immediately above the order for which one has evidence ( $> N+1$ ) has no impact on the inference process.

We now apply the above recipe to our 52 machines examples. Observe that the centre of mass of any object can be found by i) decomposing the object into disjoint parts, ii) compute the centre of mass of every part and iii) multiply each centre of mass of a part with the relative weight of this part and then iv) sum all these weighted centres of mass. We apply this well-known recipe here.

The centre of mass of  $\mathbb{C}_{\mathbb{P}_A^*}^1$ ,  $f_A^\dagger$ , assigns every  $P \in \mathbb{P}_A^*$  the same weight. We thus obtain by following the recipe

$$\begin{aligned} f_A^\dagger(P) &= \begin{cases} 0 & , \text{ if } P(\text{HEADS}) \notin [0.48, 0.52] \\ \frac{1}{0.04} & , \text{ if } P(\text{HEADS}) \in [0.48, 0.52] \end{cases} \\ f_B^\dagger(P) &= \begin{cases} 0 & , \text{ if } P(\text{HEADS}) \notin [0.54, 0.72] \\ \frac{1}{0.18} & , \text{ if } P(\text{HEADS}) \in [0.54, 0.72] \end{cases} \\ f_C^\dagger(P) &= \begin{cases} 0 & , \text{ if } P(\text{HEADS}) \notin [0.68, 0.84] \\ \frac{1}{0.16} & , \text{ if } P(\text{HEADS}) \in [0.68, 0.84] \end{cases} , \end{aligned}$$

and obtain overall

$$f_1^\dagger(P) = \begin{cases} 0 & , \text{ if } P(\text{HEADS}) \notin [0.48, 0.52] \cup [0.54, 0.84] \\ \frac{1}{52} \cdot \frac{1}{0.04} \approx 0.48 & , \text{ if } P(\text{HEADS}) \in [0.48, 0.52] \\ \frac{1}{52} \cdot \frac{1}{0.18} \approx 0.11 & , \text{ if } P(\text{HEADS}) \in [0.54, 0.68] \\ \frac{1}{52} \cdot \frac{1}{0.18} + \frac{50}{52} \cdot \frac{1}{0.16} \approx 6.12 & , \text{ if } P(\text{HEADS}) \in [0.68, 0.72] \\ \frac{50}{52} \cdot \frac{1}{0.16} \approx 6.01 & , \text{ if } P(\text{HEADS}) \in [0.68, 0.84] \end{cases} ,$$

which is plotted in Figure 5.

## 5 Discussion

### 5.1 The *right* inference process?

We have discussed three distinct ways of maximally equivocating: Williamson’s approach, and two refinements of Masterton’s approach – applying  $CM$  respectively  $CM_\infty$ . The natural question to ask is: “*So, which way is the right way?*” The answer, invariably, depends on the exact nature of our normative enterprise.

If our goal is to set subjective common-sensical probabilities as does Jeff Paris in [20]:

The probabilities assigned by Maxent (in this context) are *subjective probabilities*, quantified expressions of degrees of belief. What they are not is *estimates of objective probabilities*.  
emphasis original

then Williamson’s account (Maxent in Paris’s terms) seems to be the clear objective Bayesian frontrunner, see [21, 22, 24].

Common sensicality is cached out in terms of satisfying intuitively right (or so it is claimed) common sense principles. While each principle is, in its own right, intuitive to some degree; taken all together they are clearly not intuitive. Who in her/his right mind sets *intuitive* subjective degrees of belief equal to the unique function  $P^\dagger \in \langle \mathbb{P}^* \rangle$  which maximises  $-\sum_{\omega \in \Omega} P(\omega) \log(P(\omega))$ ? So, if rationally is construed as an explication of intuitive rationality in Carnap’s sense (see [4]), then Maxent clearly fails.

A significant number of writers, Lewis [15] for example, have defended chance-credence coordination principles that require agents to set degrees of belief equal to the chances as they know, or justifiably believe, them to be. If these

writers are right, then rational agents *do aim* at estimating chances (objective probabilities in Paris’s terminology). Maxent picks a probability function from among those left epistemically open by one’s evidence of chances, so it meets this desideratum. It is not alone in this, however; centre of mass inference processes do likewise. As we have shown, this means that inference processes—like Masterton’s—that pick a density function from among those left epistemically open by one’s evidence of chances also meet the desideratum. But Maxent,  $CM$ , and  $CM_\infty$  pick out different functions generally. Hence, the importance of the question as to which is the right inference process. Indeed, the conditions under which Maxent and  $CM$  agree are very specific and occur only rarely. We now establish what those conditions are.

**Proposition 2.** *Let  $f : \mathbb{P} \rightarrow [0, +\infty]$  be a strictly concave function which has a unique global maximum at  $P_-$ . If  $\mathbb{P}^*$  is convex, closed and  $P_- \notin \mathbb{P}^*$ , then  $\arg \sup_{P \in \mathbb{P}^*} f(P)$  contains a unique element,  $P_f^+$ , and  $P_f^+$  is an element of the boundary of  $\mathbb{P}^*$  facing  $P_-$ .*

**Proof:** For  $\mathbb{P}^*$  as above and for every point  $P \in \mathbb{P}^*$  it holds that  $f(P) < f(P_-)$ . By strict concavity of  $f$ ,  $f(\lambda P_- + (1 - \lambda)P)$  strictly decreases with decreasing  $\lambda \leq 1$  as long as  $f(\lambda P_- + (1 - \lambda)P)$  is defined, that is, as long as  $\lambda P_- + (1 - \lambda)P \in \mathbb{P}$ . For all  $Q$  in the interior of  $\mathbb{P}^*$  there exists a point  $P_Q$  in the boundary of  $\mathbb{P}^*$  ( $\mathbb{P}^*$  is closed!) which is a convex combination of  $Q$  and  $P_-$ . Hence,  $f(Q) < f(P_Q)$ . So, the maximum of  $f$  over  $\mathbb{P}^*$  cannot obtain at  $Q$ . ■

**Corollary 3.** *Let  $I$  be an inference process which picks out the probability function which maximises some strictly concave function  $f$  with a unique maximum at  $P_-$ . Let  $\mathbb{P}^*$  be convex and closed.*

- If  $P_- \in \mathbb{P}^*$ , then  $P_f^+ = P_-$ .
- If  $P_- \notin \mathbb{P}^*$ , then  $P_f^+$  is an element of the boundary of  $\mathbb{P}^*$  facing  $P_-$ .

The requirement that the unique maximum of  $f$  obtains at  $P_-$  ensures that the inference process satisfies the principle of indifference: If  $\mathbb{P}^* = \mathbb{P}$ , then  $P_f^+ = P_-$ .

**Corollary 4.** *Let  $\mathbb{P}^*$  be convex and closed.*

- If  $P_- \in \mathbb{P}^*$ , then  $P^\dagger = P_{CM_\infty}^\dagger = P_-$ .
- If  $P_- \notin \mathbb{P}^*$ , then  $P^\dagger$  and  $P_{CM_\infty}^\dagger$  are elements of the boundary of  $\mathbb{P}^*$  facing  $P_-$ .

**Proof:** It suffices to note that the two functions  $\sum_{\omega \in \Omega_I} \log(P(\omega))$  and  $H(P) = -\sum_{\omega \in \Omega} P(\omega) \log(P(\omega))$  are strictly concave on  $\mathbb{P}$  with a unique maximum at  $P_-$ . ■

**Proposition 5.** *If  $\mathbb{P}^*$  is convex and closed and if  $\mathbb{P}^*$  is of the same dimension as  $\mathbb{P}$ , then the following are equivalent*

1.  $P_{CM}^\dagger = P^\dagger$

2.  $P_{CM}^\dagger = P_-$
3.  $P_{CM}^\dagger = P_{CM_\infty}^\dagger$ .

**Proof:** 1 implies 2: We show that not 2 implies not 1. Suppose that  $P_{CM}^\dagger \neq P_-$ . There are two mutually exclusive and exhaustive cases.

Case 1  $P_- \in \mathbb{P}^*$ .

Then  $P^\dagger = P_-$ . But since  $P_- \neq P_{CM}^\dagger$  we obtain  $P^\dagger = P_- \neq P_{CM}^\dagger$ .

Case 2  $P_- \notin \mathbb{P}^*$ .

Then  $P^\dagger$  lies on the boundary of  $\mathbb{P}^*$ , on the other hand  $P_{CM}^\dagger$  is in interior point of  $\mathbb{P}^*$ .<sup>14</sup> Hence,  $P^\dagger \neq P_{CM}^\dagger$ .

2 implies 1: If  $P_{CM}^\dagger = P_-$ , then  $P_- \in \mathbb{P}^*$ . But then  $P^\dagger = P_-$ .

2 implies 3: If  $P_{CM}^\dagger = P_-$ , then  $P_- \in \mathbb{P}^*$ . But then  $P_{CM_\infty}^\dagger = P_-$ .

3 implies 2: We show that not 2 implies not 3. Suppose that  $P_{CM}^\dagger \neq P_-$ . There are two mutually exclusive and exhaustive cases.

Case 1  $P_- \in \mathbb{P}^*$ .

Then  $P_{CM_\infty}^\dagger = P_-$ . But since  $P_- \neq P_{CM}^\dagger$  we obtain  $P_{CM_\infty}^\dagger = P_- \neq P_{CM}^\dagger$ .

Case 2  $P_- \notin \mathbb{P}^*$ .

Then  $P_{CM_\infty}^\dagger$  lies on the boundary of  $\mathbb{P}^*$ , on the other hand  $P_{CM}^\dagger$  is in interior point of  $\mathbb{P}^*$ . Hence,  $P_{CM_\infty}^\dagger \neq P_{CM}^\dagger$ . ■

So,  $P^\dagger$  agrees with  $P_{CM}^\dagger$ , if and only if the centre of mass of  $\mathbb{P}^*$  is the equivocator function  $P_-$ ; this is quite a rare case.<sup>15</sup> Not only do they rarely agree, they are quite different in a great number of cases:

**Corollary 6.** *If  $\mathbb{P}^*$  is convex and closed and if  $\mathbb{P}^*$  is of the same dimension as  $\mathbb{P}$ , then  $P^\dagger$  and  $P_{CM_\infty}^\dagger$  are on the boundary of  $\mathbb{P}^*$  facing  $P_-$ .  $P_{CM}^\dagger$  is an interior point of  $\mathbb{P}^*$ .*

For a graphical illustration see Figure 4.

Convex, closed and non-empty sets  $\mathbb{P}^*$  are the paradigm examples in our setting.<sup>16</sup> In such a paradigm case with  $P_- \notin \mathbb{P}^*$  Maxent selects a point on the boundary of  $\mathbb{P}^*$ . The most basic intuition for the purposes of estimating an objective probability or at least tracking objective probabilities to the best as one can is, we would argue, to pick an interior point of  $\mathbb{P}^*$ .<sup>17</sup> We conclude that Maxent fails as an explication of everyday scientific estimation of objective probabilities because it either singles out the equivocator or a probability function on the boundary of (a convex, closed and non-empty set)  $\mathbb{P}^*$  as uniquely rational.

<sup>14</sup>Interior and exterior are here understood in the induced topology on  $\mathbb{P}^*$ . If  $\mathbb{P}^*$  has a lower dimension than  $\mathbb{P}$ , then  $P_{CM}^\dagger$  is not necessarily an interior point of  $\mathbb{P}^*$ .

<sup>15</sup>The same holds mutatis mutandis for  $P_{CM_\infty}^\dagger$ .

<sup>16</sup>Non-convex  $\mathbb{P}^*$  are notoriously hard cases. The study of note in this case which does not simply consider the convex hull of  $\mathbb{P}^*$  is [25].

<sup>17</sup>Paris & Vensovská have, of course, also found a natural way in which Maxent is uniquely rational for the purposes of estimating objective probabilities [21]. In a later paper [19, p. 275-276], Jeff Paris is somewhat more sympathetic towards  $CM$  for estimating objective probabilities.

## 5.2 The role of language

It is well-established and accepted [8, 20] that inferences have to depend in some way on the underlying language. However, the number of possible worlds depends on the agent's evidence and thus the dependence of inferences on the underlying language is rooted in the agent's epistemic state and not in the agent's whimsicality.

Language invariance on the other hand is a desideratum that can be satisfied, and indeed is satisfied, by Maxent though  $CM$  fails to satisfy this desideratum. This clearly tells against the approach we expounded here.

As we saw in Section 3, including probability densities in the agent's language allows the agent to take higher order evidence into account. In [28], Williamson expresses the worry that, in principle, agents might be required to take information on an ever higher order into account and that inference from such evidence would not eventually stabilise. A phenomenon he calls the *uncertainty escalator*. We think that higher order information, if available, ought to be taken into account and that, in every-day practise, information is only available up to a certain order. Undoubtedly, an inference process should be stable in the sense that the result of that process is the same no matter which order above that for which one has evidence one begins the inference process, but we have shown that our approach meets this desideratum. That is, we argue that our approach is not vulnerable to the uncertainty escalator for, we argue, one should ride that escalator at least as far as is warranted by one's evidence and we have shown that riding it further has no impact upon the probabilities one will infer.

Williamson is not (yet?) on the record on how exactly an agent with densities in her language sets degrees of belief. All we have to go by is his recipe to equivocate maximally (or at least sufficiently) over the most basic propositions the agent can express. We feel that it is reasonable to measure the degree of equivocation of a density distribution by (6). If this is right and Williamson acknowledges the force of expert principles such as Lewis's Principal Principle; then Williamson's agent would be forced to adopt  $CM(\mathbb{P}^*)$  as her belief function in the absence of higher order evidence, if her language includes chance hypotheses. In the same situation, but with a language that does not contain such chance hypotheses, this agent would adopt Williamson's original  $P^\dagger$  as her belief function. Clearly, Williamson's approach fails to satisfy what we have termed *chance invariance*. That is, expanding the agent's language by adding chance hypotheses while keeping the evidence base the same will alter the outcome of Williamson's version of Maxent, given some rather mild assumptions. This is not the case with our approach, though the outcome of our favoured inference process is susceptible to simple extension of the agent's language with the evidence held fixed.

Thus, considerations concerning the invariance of the results of these processes to extensions of the underlying language, far from deciding the issue, have rather produced an interesting (qualified by language dependence) scoreline of 1 : 1.

### 5.3 Non-convex evidence

As we saw in the second example (Section 2.6.2), in certain cases  $P_M^\dagger \notin \mathbb{P}^*$ . I.e., the belief function Masterton advocates adopting is sometimes ruled out by the agent's evidence. Note however, that  $P_M^\dagger \notin \mathbb{P}^*$  can only happen in cases of non-convex  $\mathbb{P}^*$ . Furthermore, for some non-convex  $\mathbb{P}^*$   $P_M^\dagger \in \mathbb{P}^*$  does hold, see Section 3.1. Finally, one should further note that this can also occur in Williamson's framework.

Overall, we feel that  $P_M^\dagger \notin \mathbb{P}^*$  for some non-convex  $\mathbb{P}^*$  is not a serious flaw, if it is a flaw at all, in Masterton's account or, for that matter, Williamson's. It is simply a result of a less than ideal evidential state.

### 5.4 Concluding remarks

In the debate as to what is the right way of equivocating between probability functions, CM and Maxent are often given prominence. Indeed, often the debate is characterised as a straight choice between the two. Aside from unfairly failing to recognise other alternatives, we judge this characterisation to be unhelpful because it fails to recognise that CM is itself a Maxent inference process. Both CM and Maxent are processes that select a probability function from those left epistemically open by one's evidence by maximising entropy, where they differ is in the entropy they are maximising. We have shown that the centre of mass of the probability functions left open by one's evidence *is* the expected function relative to the density compatible with that evidence with greatest entropy. This allows Masterton's proposal to be formulated as a centre of mass inference process, but it also allows the centre of mass of those probability functions compatible with one's evidence to be understood as the expectation relative to the evidence calibrated density with greatest entropy. Viewed in this way, we were able to generalise the centre-of-mass/Masterton's inference process to accommodate higher order evidence and to show this generalisation to be invariant to choice of starting order for such inference so long as one begins above the highest order for which one has evidence. Thus much has been gained from the discovery that the centre of mass inference process is a maximising entropy inference process of the type outlined by Masterton.

What we have not been able to achieve is a conclusive argument for *CM* against Maxent. Both inference processes have virtues and both have weaknesses. Maxent is arguably unintuitive in many situations, often picking out a probability function in the boundary of  $\mathbb{P}^*$ . Furthermore, it is not chance-invariant nor does it accommodate higher order evidence easily. *CM* is arguably more in line with scientific practice and more intuitive, and accommodates higher order evidence with aplomb. However, *CM* is not language invariant and lacks an elegant characterisation in terms of common sense principles. We have to entertain the idea that an intuitive, language invariant, chance invariant inference process which picks interior points is still to be discovered. Our recommendation in lieu of such a discovery is to adopt the Objective Bayesian inference process that best fits the prevailing circumstances, though we anticipate that this will mostly result in inference according to *CM* as presented herein.

We would like to end with quoting Jeff Paris [20, p. 6193] noting that we could not agree more:



Most of us would surely prefer modes of reasoning which we could follow blindly without being required to make much effort, ideally no effort at all. Unfortunately, Maxent is not such a paradigm; it requires us to understand the assumptions on which it is predicated and be constantly mindful of abusing them.

## References

- [1] Adamčík, M. (2014). The Information Geometry of Bregman Divergences and Some Applications in Multi-Expert Reasoning. *Entropy*, 16(12):6338–6381.
- [2] Bandyopadhyay, P. S. and Brittan, G. J. (2010). Two dogmas of strong objective bayesianism. *International Studies in the Philosophy of Science*, 24(1):45–65.
- [3] Bernardo, J. M. and Ramón (1998). An introduction to bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician*, 47(1):101–135.
- [4] Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press.
- [5] de Finetti, B. (1989). Probabilism: A critical essay on the theory of probability and on the value of science. *Erkenntnis*, 31:169–223.
- [6] Gaifman, H. (1986). A theory of higher order probabilities. In *Proceedings of the Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 275–292, California. Monterey.
- [7] Grünwald, P. D. and Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics*, 32(4):1367–1433.
- [8] Halpern, J. Y. and Koller, D. (2004). Representation Dependence in Probabilistic Inference. *Journal of Artificial Intelligence Research*, 21:319–356.
- [9] Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions On Systems Science and Cybernetics*, 4(3):227–241.
- [10] Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.
- [11] Joyce, J. M. (2010). A Defense of Imprecise Credences in Inference and Decision Making. *Philosophical Perspectives*, 24(1):281–323.
- [12] Landes, J. (2015). Probabilism, Entropies and Strictly Proper Scoring Rules. *International Journal of Approximate Reasoning*, 63:1–21.
- [13] Landes, J. and Williamson, J. (2013). Objective Bayesianism and the maximum entropy principle. *Entropy*, 15(9):3528–3591.

- [14] Lawry, J. and Wilmers, G. M. (1994). An axiomatic approach to systems of prior distributions in inexact reasoning. In Masuch, M. and Pólos, L., editors, *Knowledge Representation and Reasoning Under Uncertainty*, volume 808 of *LNCS*, pages 81–89. Springer.
- [15] Lewis, D. (1980). A Subjectivist’s Guide to Objective Chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*, volume 2, chapter 13, pages 263–293. Berkeley University Press.
- [16] Masterton, G. (2015). Equivocation for the Objective Bayesian. *Erkenntnis*, 80:403–432.
- [17] Miller, D. (1966). A paradox of information. *British Journal for the Philosophy of Science*, 17(1):59–61.
- [18] Paris, J. B. (1994). *The Uncertain Reasoner’s Companion*. Cambridge University Press.
- [19] Paris, J. B. (2005). On filling-in missing conditional probabilities in causal networks. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 13(3):263–280.
- [20] Paris, J. B. (2014). What You See Is What You Get. *Entropy*, 16(11):6186–6194.
- [21] Paris, J. B. and Vencovská, A. (1989). On the applicability of maximum entropy to inexact reasoning. *International Journal of Approximate Reasoning*, 3(1):1–34.
- [22] Paris, J. B. and Vencovská, A. (1990). A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, 4(3):183–223.
- [23] Paris, J. B. and Vencovská, A. (1992). A method for updating that justifies minimum cross entropy. *International Journal of Approximate Reasoning*, 7(1–2):1 – 18.
- [24] Paris, J. B. and Vencovská, A. (1997). In Defense of the Maximum Entropy Inference Process. *International Journal of Approximate Reasoning*, 17(1):77–103.
- [25] Paris, J. B. and Vencovská, A. (2001). Common sense and stochastic independence. In Corfield, D. and Williamson, J., editors, *Foundations of Bayesianism*, pages 203–240. Kluwer.
- [26] van Fraassen, B. (1980). *The Scientific Image*. Clarendon Press, Oxford.
- [27] Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford University Press.
- [28] Williamson, J. (2014). How Uncertain Do We Need to Be? *Erkenntnis*, 79(6):1249–1271.
- [29] Wilmers, G. (2015). A Foundational Approach to Generalising the Maximum Entropy Inference Process to the Multi-Agent Context. *Entropy*, 17(2):594–645.