Bertrand's Paradox and the Maximum Entropy Principle

Nicholas Shackel and Darrell Rowbottom[1]

Abstract: An important suggestion of objective Bayesians is that the maximum entropy principle can replace a principle which is known to get into paradoxical difficulties: the principle of indifference. No one has previously determined whether the maximum entropy principle is better able to solve Bertrand's chord paradox than the principle of indifference. In this paper I show that it is not. Additionally, the course of the analysis brings to light a new paradox of the chords that is unique to the maximum entropy principle.

1. Introduction

An important suggestion of objective Bayesians is that the maximum entropy principle can replace a principle which is known to get into paradoxical difficulties: the principle of indifference. The maximum entropy principle can be considered as a refinement and generalisation of the principle of indifference: they agree on cases to which the latter principle applies whilst the former can yet be used where the latter has no purchase.

Bertrand constructed his chord paradox to pose a challenge to the principle of indifference by showing that it led to inconsistent probabilities. Despite an extensive literature aimed at resolution, Bertrand's chord paradox continues to undermine the principle of indifference (Shackel 2007).

There is no such literature examining whether the maximum entropy principle is able to resolve Bertrand's paradox. Indeed, there is no literature even showing how the maximum entropy principle can get a purchase on the paradox. In this paper I show that even under the most favourable assumptions allowing for that purchase, Bertrand's chord paradox undermines the maximum entropy principle. Additionally, the course of the analysis brings to light a new paradox of the chords that is unique to the maximum entropy principle.

2. The maximum entropy principle

Salmon's Ascertainability criterion requires 'that there be some method by which, in principle at least, we can ascertain values of probabilities' (Salmon 1967:64). Classical probabilists such as Bernoulli (1713) and Laplace (1814/1995)[2] and logical probabilists such as Keynes (1921/2014) have used as their method the principle of indifference:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favourable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability (Laplace 1814/1995:6-7)

---

[1] Author's contributions: Shackel 90%, Rowbottom 10%. In memoriam Michael Clark 1940-2019, who shared our delight in the philosophical problems posed by paradoxes.

[2] Mathematicians may think of classical probabilism as identifying probability with a normalized measure as axiomatized by Kolmogorov, but this is not what is meant philosophically.

> if there is no known reason for predicating of our subject one rather than
> another of several alternatives, then relative to such knowledge… these
> alternatives have an equal probability. (1921/2014: 45).

The mathematical theory of probability alone tells me nothing of the probability of heads, but because I have no reason to expect either side of the coin, heads has the probability of one half.

Unfortunately the principle of indifference falls afoul of a number of paradoxes, most famous of which is Bertrand's chord paradox. When the principle of indifference is applied to the paradox it results in three different probabilities for one and the same event. This non-uniqueness was taken to be a refutation of classical probabilism (Bertrand's own target in constructing the paradox) and has also been taken to be a refutation of logical probabilism.

Objective Bayesians such as Jaynes (2003), Rosenkrantz (1977) and Williamson (2010) have suggested that the maximum entropy principle can replace the principle of indifference.

> The maximum entropy principle…the prior probability assignment should
> be the one with the maximum entropy consistent with the prior knowledge.
> (Jaynes 1968:229)

Jaynes proposes this because the maximum entropy probability function is 'maximally non-committal with regard to missing information' (Jaynes 1957a:620) and 'the most unbiased representation of our knowledge' (Jaynes 1957a:171). Jaynes reiterates this thought with slight variation through his papers (Jaynes 1957a, b, 1963, 1968, 1985). In his book (2003) he explicitly joins it up with Shannon's development of entropy in information theory. 1814/1995

The reason entropy helps Jaynes here is the thought that bias is a kind of unwarranted lack of uncertainty, so being maximally unbiased would be maximising uncertainty, and Shannon's suggestion (1948) that entropy can be understood as a measure of uncertainty.[3] Here are some platitudes deploying our pre-theoretical concepts of uncertainty and probability:

1. uncertainty depends on how likely events are and is entirely absent if all but one have no chance at all,
2. we are most uncertain when possibilities are equally likely,
3. uncertainty does not change if we add events with no chance of happening and
4. adding new events with some chance of happening makes us more uncertain in a way that depends on the uncertainty of the events added. [4]

Shannon gave initial proofs that his entropy satisfies mathematical requirements we can see as capturing much of these platitudes (1948:10 & 28). Khinchin shows more rigorously (Khinchin 1957:9 Thm 1) that Shannon entropy satisfies Khinchin's three axioms for uncertainty,[5] axioms that we might reasonably take to be the mathematical

---

[3] Clearly nothing turns on whether we formulate this in terms of certainty instead of uncertainty. It is done this way because Shannon entropy increases with increasing uncertainty, i.e. increases with decreasing certainty.

[4] The use of 'chance' here is colloquial rather than a deployment of the term stipulated by philosophers to mean objective probability.

[5] '1. For given n and for $\sum_{k=1}^{n} p_k = 1$, the function $H(p_1, p_2, \ldots p_n)$ takes its largest value for $p = 1/n$ ($k = 1, 2, \ldots n$). 2. $H(AB) = H(A) + H(B)$. 3. $H(p_1, p_2, \ldots p_n, 0) = H(p_1, p_2, \ldots p_n)$.' (Khinchin 1957:9)

equivalents of the first three platitudes above (Khinchin 1957:9 and Beck and Schlögl 1993:47).

So Shannon's entropy is a reasonable measure of our intuitive notion of uncertainty. Taking the probability function compatible with information possessed that otherwise *maximizes* uncertainty is a way to *minimize* bias in representing our knowledge. Hence using the maximum entropy principle seems to be a way of satisfying the Ascertainability criterion that fits with Jaynes' claims for the principle. In capturing the second platitude the maximum entropy principle can be regarded as a generalization of the principle of indifference.

The maximum entropy principle has been applied to some of the paradoxes that trouble the principle of indifference and it has been claimed to resolve some of them.[6] So it may well appear that the maximum entropy principle is a good prospect for solving Bertrand's paradox (indeed, there is an impression that Jaynes (1973) did exactly that, see more below). In fact, and perhaps surprisingly given the extensive literature on Bertrand's paradox and the principle of indifference, nowhere in the literature has there been an analysis of applying the maximum entropy principle to Bertrand's paradox. By an analysis, I mean showing whether the mathematics of entropy, when applied to Bertrand's paradox, determines a unique probability function with maximum entropy. I demonstrate here that it does not.

Although this result may have been suspected by some, it has not previously been proved. To put it crudely, absent this paper, for all we know the maximum entropy principle might yet solve Bertrand's chord paradox. This paper proves it does not. Moreover, the course of the analysis exhibits a further problem for the maximum entropy principle: I present a new paradox of the chords that arises only for the maximum entropy principle.

Having proved all this, the work of the paper is done. We have, of course, brought into view a further question, namely, what kind of a problem is non-uniqueness for proponents of the maximum entropy principle? Before saying more on that, we need to dispel the impression that prior literature has analysed the application of the maximum entropy principle to Bertrand's paradox.

The source for this impression is the existence of the paper on Bertrand's paradox by the original proponent of the maximum entropy principle, namely, Jaynes (1973). Rosenkrantz explicitly treats Bertrand's paradox in his book (1977:73), but whilst the context may give the impression that he is applying the maximum entropy principle, in fact he is expounding Jaynes' paper.

Jaynes' abstract for his paper is misleading: it speaks of the maximum entropy principle but the paper itself does not apply the maximum entropy principle to Bertrand's paradox. Instead it makes use of the well-posed/ill-posed distinction and then applies Jaynes' symmetry invariance principle and the principle of indifference. The question of what it is for something to be ill-posed is analysed in Shackel (Shackel 2007), where it is shown that Jaynes' solution does not work. In our opinion Shackel has refuted Jaynes' paper.

$H(p_1,p_2,…p_n)$ is defined to be 'a quantity that measures the amount of uncertainty' (Khinchin 1957:3).

[6] The book paradox, for example. Under the principle of indifference, probability of book being colourful = ½ or 7/9 (or…) depending on how fine grained one takes the ignorance (colourful versus not-colourful, one of seven colours versus black or white,…). Paris and Vencovská 1997 claim that correct application of the maximum entropy principle defeats this paradox. For a recent summary of Paris's solutions see Paris 2014.

Some may claim that Jaynes' view of the maximum entropy principle as a generalization of the principle of indifference means there is a sense in which Jaynes' paper 'applies' the maximum entropy principle to Bertrand's paradox by applying the principle of indifference to Bertrand's paradox. This thought is tempting[7] but too quick when we are talking about infinite event spaces, although that will only be evident after we have plowed through the work herein. What we can do for now is to show that this thought is a distraction.

First, if Jaynes was in some sense 'applying' the maximum entropy principle in his paper, Shackel's refutation means that it was not done successfully. Second, the mathematics of entropy are not actually used in Jaynes' paper.[8] No equations for the entropies of any probability functions are given and nowhere does the paper make any attempt to show that the probability function derived is the one with maximum entropy. Neither entropy itself nor the criterion of having maximum entropy do any work in the paper. Consequently, whilst Shackel's refutation of Jaynes' paper shows that this special sense of 'applying' the maximum entropy principle is unsuccessful, it does not show that the maximum entropy principle itself cannot solve Bertrand's paradox. For all we know, then, were we to actually use the mathematics of entropy, the maximum entropy principle might yet solve Bertrand's paradox. It is only the work herein that proves it does not.

Insofar as Williamson discusses Bertrand's paradox and the maximum entropy principle (2010:152), he merely mentions it, mentions Jaynes' (1973) paper and perhaps endorses it. He makes no attempt to apply the maximum entropy principle to Bertrand's paradox nor does he make any statement on whether doing so would result in unique probabilities. Instead he moves on immediately to indicating that he is untroubled by non-uniqueness in general. That, however, takes us to the further question.

Now there is no doubt that this is an important question in the philosophy of probability. It should, of course, be remembered that precisely this failure of uniqueness for Bertrand's paradox on the part of the principle of indifference was taken to be a refutation of classical and logical probabilism—the reason being that non-uniqueness, absent additional and substantive philosophical theory, is a polite term for inconsistency.[9] That is why Jaynes originally intended that the maximum entropy principle should result in unique probabilities (see Jaynes 1968). So plainly there is a threat to objective Bayesianism in our result. Some modern objective Bayesians such as Williamson, on the other hand, claim that non-uniqueness is not a problem, but whether that is mere whistling past the graveyard depends on the answer to the further question.

Addressing this question would involve considering whether and what substantive philosophical theory has been offered to avoid non-uniqueness being inconsistency, what kind of restriction on event spaces that would amount to, what theoretical costs are acceptable in that restriction, where all this would have to be examined in the light of the broader epistemic principles that ground our philosophical interest in probability. The further question is hence part of a general question in epistemology over uniqueness

---

[7] Tempting because of the second platitude above and because for finite event spaces it is easy to prove that the uniform probability function is the function with maximum entropy.

[8] An examination of the appendix herein will show what would have been necessary for the mathematics of entropy to have been used.

[9] It may appear that we need to prove the non-unique probabilities for Bertrand's paradox arise from equivalent descriptions of the same random phenomenon to move from non-uniqueness to inconsistency. This is a version of what Shackel calls the distinction strategy, whose failure as an avoidance of inconsistency when used in this way for Bertrand's paradox is shown in Shackel 2007:§10.

versus permissivism.[10] For all their whistling, by the very name by which they distinguish themselves from subjective Bayesians, the objective Bayesians place themselves on the side of uniqueness against permissivism and, furthermore, to be on that side was indeed the intention of Jaynes in so naming it.[11]

Unfortunately, none of this can be addressed briefly. It merits its own full treatment (see Shackel MS-b). So for reasons of space we have had to eschew the further question that arises from our results for proponents of the maximum entropy principle.

Finally, the interaction of the maximum entropy principle and Bertrand's paradox has not been previously examined by philosophers perhaps in part because entropy can involve highly abstract mathematics. On the other hand, some proponents of the maximum entropy principle are mathematicians and physicists who are uninterested in or intolerant of the philosophical issues, and consequently, some of its specialist literature does the maths whilst discarding awkward philosophical problems as mere matters of convention, or alternatively, burying them in the maths. This is off-putting for philosophers and I have sought to do the opposite. It turns out that although we need to take the output of the relevant mathematics as premises, the central philosophical issues do not depend much on why those premises are true. For this reason, whilst I have included the mathematics sufficient to the latter question in an appendix, I have been able to articulate the philosophical import of the problems without requiring the reader to plough through that mathematics.

3. Bertrand's Paradox[12]

Bertrand posed his paradox thus:

> We trace *at random* a chord in a circle. What is the probability that it would be [longer] [13]than the side of the inscribed equilateral triangle? (Bertrand 1888: 4)
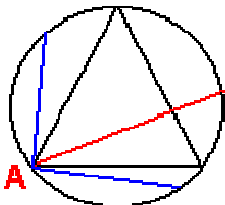
As is well known, applying the principle of indifference in three different ways gives three different answers for the probability of longer (as we shall hereafter call the probability of getting a chord longer than the side of the inscribed equilateral triangle).[14] Of course, if we specify a particular way of choosing chords at random we can avoid this, but doing so is merely an evasion of the generality of the problem posed by Bertrand's paradox.[15]

---

[10] E.g. see White 2005; Greco and Hedden 2016

[11] See Jaynes 1968

[12] This section draws in part on material from Shackel 2007.

[13] Bertrand himself said shorter, but the literature has used longer ever since so we stick with this.

[14] The probability of longer is ¼, 1/3 and ½. For reasons of space I do not here show the application of the principle of indifference to Bertrand's paradox. For details see Clark 2002:18 or Shackel 2007

[15] For in depth discussion on this point, see Shackel 2007:164-9, and also van Fraassen 1989:305.

Prior work (Shackel 2007 sections 5-8) shows that for our analysis we need only the most fundamental features of the standard measure theoretic formulation of probability. For reasons of space we offer here a summary rather than a recapitulation of that prior work. A probability space is an ordered triple $\langle X, \Sigma, P \rangle$, where X is the space of 'events', Σ is a σ-algebra on X and P is a measure on Σ for which $P(X) = 1$.[16] Being such a measure is sufficient for satisfying Kolmogorov's original probability axioms (Capinski and Kopp 2004, 46 Remark 2.33).[17]

Let C be the set of chords with which we are concerned. We must not let C's close association with the Euclidean plane, $\mathbb{R}^2$, blind us to the fact that C is not a subset of $\mathbb{R}^2$. Rather, because each chord is a subset of $\mathbb{R}^2$, C is a subset of the power set of $\mathbb{R}^2$, $\mathbb{P}(\mathbb{R}^2)$. σ-algebras of intervals of $\mathbb{R}^2$ and the measures on $\mathbb{R}^2$ are not σ-algebras and measures of C. Unlike $\mathbb{R}^2$ and other cases we are used to dealing with, there is no 'natural' prior measure[18] on C: C lacks a measure of its own. Consequently to get a probability measure on C we need to use Shackel's

> Theorem of Induced Σ and μ: Given a set, Y, with a σ-algebra, A, and a measure, m, we can use a suitable[19] function $f{:}X{\rightarrow}Y$ (a bijection is sufficient but not necessary) to induce a measure on the set X. We define Σ to be the set of pre-images of members of A, and define the measure under μ of an element in Σ to be the measure under m of its image set in A.[20][21](Shackel 2007:160)

We can then use the induced sigma algebra and measure to construct the probability function by setting $P(x) = \mu(x)/\mu(X)$ for all x in Σ whenever $\mu(X)$ is bounded.

So we must make use of functions from C into measurable sets, and that is what Bertrand is seeking to do when, in his original presentation, he equates measures on C with measures on $\mathbb{R}$ in two cases and a measure on $\mathbb{R}^2$ in the third. But of course, there

---

[16] We shall continue to speak in terms of events, but X can just as well be a mathematical space or set of possible worlds or propositions or outcomes, according to taste. A σ-algebra is a set, A, of subsets of a set, S, (so $A \subseteq \mathbb{P}(S)$) that contains S and $\varnothing$, and is closed under complementation and countable union. A measure for A is a non-negative function $\mu{:}A{\mapsto}\mathbb{R}$ such that $\mu(\varnothing) = 0$ and μ is countably additive.

[17] So far as the concerns of this paper go, alternate axiomatizations by subjectivists such as Ramsey, de Finnetti and Savage make no difference. See Fishburn: 'The theory of subjective probability attempts to make precise the connection between coherent dispositions toward uncertainty and quantitative probability as axiomatized by Kolmogorov' (Fishburn 1986:353). Ramsey, for example, proves that his definition of rational degrees of belief as betting odds satisfies the standard measure theoretic axioms for his sigma algebra of beliefs.

[18] The contrast is with, for example, the 'naturalness' of taking the measure of a line segment to be the distance between its ends.

[19] For anyone stumped by 'we can use a suitable function $f{:}X{\rightarrow}Y$ to induce', read 'a function $f{:}X{\rightarrow}Y$ induces'.

[20] This theorem is easily proved by elementary methods and more concisely from the fact that a map between sets yields a morphism of set systems (See Jost 2015:Lemma 4.3.1). Then, obviously, $\mu(X){\neq}0$ and showing countable additivity is fairly trivial. For a simple example, suppose $f{:}X{\rightarrow}Y$ is a bijection. Let $f(S) = \{y{\in}Y: y = f(x)$ for some $x{\in}S\}$; then S is the pre-image of $f(S)$ and we define Σ and μ thus: $S{\in}\Sigma$ iff $f(S){\in}A$ and $\mu(S){=}m(f(S))$.

[21] Use of this theorem need not be explicit. Any probability function assigned to C will determine a class of functions each of which can perform the role of the function in the theorem. See Shackel 2007:160-1 for how Bertrand is making implicit use of the theorem.

are infinitely many such functions, and it is clear that for any $x \in [0,1]$ we can induce a measure that gives us $P(\text{longer}) = x$.[22]

4. The maximum entropy principle applied to Bertrand's Paradox

The essential problem of Bertrand's paradox for the principle of indifference is that the set of chords, C, lacks a measure of its own by which to define equi-probability for equal ignorance over C; instead there is only an infinity of measures we might induce on C and no principled way of picking among them. This lack of a measure of its own and the necessity to use the Theorem of Induced Σ and μ is also why Bertrand's paradox is richer than other paradoxes and, because there are many such sets, the source of its significance as an example of a pervasive problem.[23]

The apparent virtue of applying the maximum entropy principle is that, rather than *ignoring* the infinitude of inducible probability functions (between which nothing we know gives us reason to discriminate), taking the probability function with maximum entropy amounts in this case to exercising what we might call meta-indifference over that infinitude. Taking the one with maximum entropy is a way of choosing the maximally ignorant one, the least biased of that infinitude.

For this to work requires that

1. there is a correct measure of entropy and it can be applied to Bertrand's paradox
2. each probability function in the set of probability functions inducible on C by the Theorem of Induced Σ and μ has a unique entropy,
3. we can identify at least one function with maximum entropy,
4. and if there is more than one function with maximum entropy they all agree on the probability of the longer chord.

There are some significant technical difficulties that stand in the way of the first two of these, and some of those difficulties, if not resolved, do reinstate the paradox. However, I have traced a path through the difficulties, making use of Shannon's differential entropy, that a proponent of the maximum entropy principle may reasonably take. For that reason, and to avoid obstructing the flow of argument, I have described those difficulties and a path through them in the appendix to this paper.

As I shall now show, Bertrand's paradox exhibits serious obstructions that lie in the way of fulfilling the other conditions (3 and 4 of our list). I now demonstrate three

---

[22] Proof: for any $x \in [0,1]$, in the Theorem of Induced Σ and μ we can use any function from C to [0,1] for which the image of {longer chords} is $[0, x]$ and the image of {not-longer chords} is $(x, 1]$.

[23] A recent paper claims to 'interpret' Bertrand's chord paradox in a way that 'does not undermine the principle of indifference' (Gyenis and Rédei 2015:349). The relation of Gyenis and Rédei's treatment to the analysis here is rather complex, including substantive disagreements about the nature of the principle of indifference and how Bertrand's paradox does its damage, and obscurity in the relation of Bertrand's actual chord paradox to what Gyenis and Rédei call their 'General Bertrand's paradox: Let (X, S, p_H) and (X', S', p'_H) be probability spaces with compact topological groups X and X' having an infinite number of elements and p_H, p'_H being the respective Haar measures on the Borel σ-algebras S and S' of X, and X'.' (2015:359). Gyenis and Rédei do not explain the relation. With the tool used herein we can get to such a pair of probability spaces by two applications of the Theorem of Induced Σ and μ from a pair of compact topological groups and define correlate compact topologies on C by the same function that induces the measure. But now we are well down stream of Bertrand's paradox. The status of this interpretation and the other issues that arise are the subject of an entire paper, Shackel MS-a, so for reasons of space I do not attempt to address them further here.

obstructions, each of which, if correct, shows that the maximum entropy principle fails to determine a unique probability for being longer.

5. Unknowable maximum entropy probability functions are not acceptable.

Bertrand's paradox brings into view the danger that spaces without measures of their own pose to knowing the maximum entropy probability function. If there are cases in which we cannot know, then the principle fails to satisfy Salmon's ascertainability criterion. We do not know the maximum entropy probability function over the set of chords and, as I shall now show, we have good reason to think we cannot know it.

Many integrals are not analytically solvable—call these obstinate integrals.[24] If the integrals are obstinate, we don't know if the integrals are numerically tractable (approximately solvable in reasonable time using numerical methods), and we do know there are serious difficulties in evaluating continuous entropies even just for $\mathbb{R}^n$.[25] This is a neglected issue because physicists are used to finding methodological tweaks to get round obstinacy and their empirical focus means they do not attend to the fully general problem that concerns us as philosophers of probability.

One route to knowing the maximum entropy probability function would be to evaluate entropies. Unfortunately, this is blocked by obstinate integrals leading to numerical intractability. The necessity of using the Theorem of Induced Σ and μ means we know that there will be functions from C, onto codomains with measures giving rise to obstinate integrals, that induce measures on C resulting in obstinate integrals for defining entropy for probability functions over C. Since we cannot solve analytically for their entropy, solving for the probability function with maximum entropy over C may require numerical methods. However, many such will not be numerically tractable and that would mean we won't be able to know the maximum entropy probability function in this way.

Numerical intractability, it might be objected, poses a merely practical problem. In principle, numerical evaluation is possible and the intractability is just a matter of it taking a very long time to do the figuring. This, it must be conceded, is certainly true of each individual obstinate integral. Unfortunately, since there are infinitely many such integrals, this does not offer an in principle defence of knowability. The problem is that to determine the maximum entropy principle in this way would require performing infinitely many numerical evaluations and this we cannot do.

The other route we might take to knowing the maximum entropy probability function would be having a constructive existence proof of which one it is that does not depend on evaluating entropies. The necessity of using the Theorem of Induced Σ and μ means the set of probability functions over which we seek the one with maximum entropy, PF_C, is both enormous[26] and essentially unrestricted.[27] The size of PF_C means, absent relevant

---

[24] Roughly, lack a solution expressible as a formula not involving differential or integral equations, power series or limits. If a solution is expressible in closed form it is possible to evaluate in finitely many steps and for that reason called tractable. Analytically solvable is slightly wider than closed form solvable by allowing some special functions such as the gamma function and some infinite expressions.

[25] See Pearl (1988 chapter 9) who objects to the maximum entropy principle on the grounds of computational intractability.

[26] Take a probability space (C, S, P) where P is an induced probability function on C for which the subset of the sigma algebra, S, having non-zero probabilities is continuum sized, with continuum many distinct probabilities. Any permutation of S corresponds to a distinct probability function on C. Any such distinct probability function, P', on C, can be induced on C using the Theorem of Induced Σ and μ (take any continuum sized set, X, and use a bijection $f:X \to C$ to induce from the given P' a measure on X, (X, T, M), and now, using f^{-1} to induce a probability function on C from

proofs, there is no reason to expect such a constructive proof would be available. Since it is unrestricted, that there are some sets of probability functions for which such constructive proofs are available, which sets will be subsets of PF_C, tells us very little. There are, after all, also sets of probability functions for which there are no such constructive proofs and they, too, will be subsets of PF_C. So consider the union of a set for which there is a constructive proof and a set for which there isn't, and for which the entropies of the first are all less than those of the second. There will be no constructive proof for such a union and the union will be a subset of PF_C. This counts against such a constructive proof being available for PF_C.

It is true that there may be specific empirical situations whose modelling sufficiently restricts the set of probability functions induced on C to one for which no entropy is defined by an obstinate integral or one for which there is a non-evaluative constructive proof. But that is not the general case.

Although these several points do not constitute a conclusive proof that (if there is one) the unique maximum entropy probability function for Bertrand's paradox is unknowable, they are good reason to think that it is not. We need to be clear why not knowing or being unable to know amounts to a failure. After all, many problems in real analysis are intractable. The nature of the problem here depends on whether the maximum entropy principle is taken to be a purely metaphysical principle or whether it is in part or wholly an epistemic principle. [28]

If the maximum entropy principle is taken to be a purely metaphysical principle, in the sense that it *determines* a single correct probability function for Bertrand's paradox, whether or not we are able to *know* which it is, then perhaps this is not a failure. There just has to *be* a unique function with maximum entropy. Nevertheless, in this case the proponents of the principle owe us a general existence proof since otherwise they are merely claiming that the principle will solve the paradox. No such proof has ever been offered for Bertrand's paradox. Shortly I show that there is a function for Bertrand's paradox with maximum entropy, but not in a way that will help proponents.

If, however, the principle is in part or purely an epistemic principle then this is a serious problem. The role of the principle in, for example, objective Bayesianism is epistemic. Indeed, among Jaynes' philosophical reasons for advancing the principle was to solve the problem of justifying applying statistical mechanics to a deterministic physical system. Jaynes considered that standard explanations of statistical mechanics involved objectionable obscurities whereas:

> [the] independence [of Jaynes' derivation of the Boltzmann distribution] from difficult and dubious physical assumptions like ergodicity, and its avoidance of fictions like 'virtual ensemble' have impressed many students of statistical mechanics and inductive logic. (Shimony 1985:38)

Jaynes gains this independence because he bases the derivation on an application of the maximum entropy principle. This allows probabilistic properties of statistical mechanics

---

(X, T, M) will induce the given probability function, P'. The cardinality of the set of such permutations $= 2^{|S|} = \text{Beth}_2 > $ cardinality of the continuum$= \text{Beth}_1$. Hence the set of all probability functions inducible on C is at least as big. (Relation to Aleph cardinals: $\text{Beth}_n \geq \text{Aleph}_n$. The general continuum hypothesis implies equality.)

[27] Since $|C| = |\text{continuum}|$, for any continuum sized space, X, there is a bijection from it to C and so for any bounded measure on X that bijection can be used with that measure in the Theorem of Induced Σ and μ.

[28] For discussion on the same point as it arises for the principle of indifference, see Shackel 2007:161-2

to be understood in epistemic terms rather than requiring them to be physically real. Consequently, statistical mechanics is not a purely physical theory but is instead an epistemic theory of our knowledge of thermodynamic systems. Clearly all this requires taking the maximum entropy principle to be an essentially epistemic principle.

Finally, the very fact that the principle is supposed to address the Ascertainability criterion shows it to be at least in part an epistemic principle. The whole point of the principle is to produce determinate probabilities when we do not know, without restriction on the extent of our ignorance. Even if the principle succeeds in determining that there are such probabilities, if we cannot know what they are because we cannot know which probability function has maximum entropy, then it has failed at the point of that point, namely, to be the method by which we can satisfy Salmon's Ascertainability criterion.

Some Bayesians countenance the axioms of probability as regulative ideals that we cannot know whether we meet because our finitude entails lack of logical omniscience and probabilistic completeness. It might then be argued that the maximum entropy principle could function similarly despite the unknowability of the function with maximum entropy (provided we had an existence proof). The question then would be how well it could function as such. It might give some local direction when one function had higher entropy than another, but we shall be seeing below that in general there will be too many functions with the same entropy and sometimes unbounded entropies for that to help. I shall see below that Bertrand's paradox is a case in point and hence the principle does not solve the paradox even when taken as a regulative ideal.

6. Either Bertrand's paradox returns or we have a new paradox of chords

There is an important difference between entropy for discrete and continuous event spaces: entropy can be unbounded for continuous distributions (more loosely put, can be infinite). We seek the maximum entropy over *all* probability functions that could be induced on *C*. Since entropy can be unbounded for continuous distributions there will be probability functions of that type that can be induced on *C*. Indeed, there will be infinitely many probability functions of that type that can be induced on *C*. The fact that they might not be well motivated geometrically is beside the point of the strategy of going for the maximum entropy probability function. Therefore there will be no single maximum entropy probability function. So for the maximum entropy principle to solve Bertrand's paradox would require all the probability functions for which entropy is unbounded to agree on the probability of longer. Given the lack of structure in *C*, provided there is just one such for which the probability of longer $> \frac{1}{2}$ there will be another that is its dual[29] giving the probability of longer $< \frac{1}{2}$. That would immediately give rise to Bertrand's paradox. So for this strategy to work would require that all the probability functions, *P*, for which entropy is unbounded give the probability of longer, $P(L) = \frac{1}{2}$.

Now consider the following events whose disjunction is the same as *not being longer*: b*eing less than half the length of the inscribed equilateral triangle*; *being between half the length and the length of the inscribed equilateral triangle*. The first partitions *C* into sets *H* and ¬*H* and the second into *G* and ¬*G*. The duality explained in footnote 29 applies to these sets also and hence so does the reasoning just given about the event of *being longer*. Hence, on pain of the immediate recurrence of Bertrand's paradox, all the probability

---

[29] Since longer and not longer partition *C* into sets *L* and ¬*L*, a suitable dual will be got from the function, *f*, inducing probability of longer$> \frac{1}{2}$ by way of the Theorem of Induced Σ and μ (Shackel 2007:160), by taking any function, *g*, such that $g(L) = X$ iff $f(\neg L) = X$. So in this case the measure of $f(L) > \frac{1}{2} \times \mu(C)$ iff the measure of $g(L) < \frac{1}{2} \times \mu(C)$ (because $\mu(C) = \mu(f(L)) + \mu(f(\neg L)) = \mu(f(L)) + \mu(g(L))$).

functions, P, for which entropy is unbounded must give $P(H) = \frac{1}{2}$ and $P(G) = \frac{1}{2}$. Now we have a new paradox of the chords:

$\frac{1}{2} = P(L)$

$= P(\neg L)$

$= P(H \text{ or } G)$

$= P(H) + P(G)$ (by additivity of measures, since $H \cap G = \varnothing$)

$= \frac{1}{2} + \frac{1}{2} = 1$

So, on the assumption that, despite unbounded entropies, there is a solution to Bertrand's paradox furnished by the maximum entropy principle, a new chord paradox appears.

The only way out of this is to exclude probability functions with unbounded entropies. Yet the reason grounding the maximum entropy principle is that by rejecting those with less than maximum entropy we reject all probability functions that are more biased than is warranted by our state of knowledge. Consequently it is very difficult to see how we could justify excluding such functions.

I am therefore confident that this challenge is correct and conclusive. There is no unique maximum entropy probability function for the set of chords, but infinitely many which have unbounded entropies and are therefore equally maximal. Either they do not agree on a unique probability of longer or they do. If they don't we have Bertrand's paradox again, but if they do then we end up in the new paradox.

7. Coda: collapse into the original Bertrand's paradox

Since I see no prospect for a principled exclusion of probability functions with unbounded entropies, I believe that at this point I have established the result of the paper. We continue with a coda only for the sake of demonstrating that *even if* someone were to come up with such a principled exclusion, its effect would not save the maximum entropy principle from Bertrand's paradox but result in it collapsing into a version of the original Bertrand's paradox.

So now we exclude the induced probability functions on C with unbounded entropies. With this exclusion another problem comes into view: that although all probability functions left have finite entropies the set of entropies need not have a maximum. So we shall also have to assume that problem away. We shall now see quite directly that there will still be more than one probability function with maximum entropy and those probability functions will disagree on the probability of the longer chord. Furthermore, despite the apparent simplicity of the route now taken, the tactics applied by proponents of the maximum entropy principle to simpler paradoxes cannot get a grip just because of the richness of Bertrand's paradox. Any such attempt will amount to a covert restriction rather than comprehension of the case and be thereby an evasion of the general problem posed by Bertrand's paradox.

Shannon entropy satisfies Khinchin's second axiom, that entropy is maximum iff the probability function is uniform.[30] This is no accident. Khinchin's second axiom effectively captures one of the central simple truths about uncertainty, that we are most uncertain when possibilities are equally likely. Consequently, if a measure of entropy does *not* satisfy Khinchin's second axiom then that brings into question the use of that measure for

---

[30] Shannon asserts the truth of this when the underlying event space has bounded volume but does not address the completely general case. Shannon 1948:35

applying the maximum entropy principle. It means that the argument I gave above defending the maximum entropy principle on the grounds of platitudes about uncertainty would not apply to that measure. It would mean that the measure fails to satisfy Williamson's equivocation norm (Williamson 2010:28). It would be contrary to what Jaynes requires when he says the maximum entropy principle reduces to 'the principle of insufficient reason…in case no information is given except enumeration of the possibilities' (Jaynes 1957a:623), since to do this the principle must give the uniform distribution in the case of complete ignorance.

From this it follows (under the assumptions made at the beginning of this section) that the maximum entropy probability functions on C are uniform distributions when not constrained by further information. In Bertrand's paradox we are not constrained. Uniform distributions on C are defined, via the Theorem of Induced Σ and μ, by uniform distributions on the inducing space. From the original version of Bertrand's Paradox we know of two inducing spaces whose uniform distributions induce uniform distributions on C, say P_1 and P_2.[31] By the measure satisfying Khinchin's second axiom, entropies of both P_1 and P_2 are equal and maximum. P_1 and P_2 give contrary probabilities for the longer chord. Therefore Bertrand's Paradox returns under the maximum entropy principle.

8. Conclusion

There are some assumptions about the maximum entropy principle in the air: that whether it can solve Bertrand's chord paradox is not interesting, that it is not a surprise if it can't, that Jaynes used it to solve the paradox and in any case the need for uniqueness must be proven before anyone should care about this so any analysis of the question lacks a target. No doubt such rhetoric is useful for those who wish to whistle past graveyards. I, however, contend that these assumptions are erroneous.

The relation of the principle of indifference and Bertrand's paradox has given rise to an extensive analytical literature. There is not a single paper that has conducted an analysis of the relation of the maximum entropy principle and Bertrand's paradox. Assumptions about the interest or surprise in the absence of such analysis are prejudices. Jaynes did not even attempt to use the principle to solve the paradox, but did want uniqueness. The suggestion that the large question of uniqueness must be settled before anyone should care about such analysis is a mere evasion.

In this paper I have rectified the error. I conduct the needed analysis and prove that the maximum entropy principle cannot solve Bertrand's chord paradox. The wider significance of this is part of the entire stream of debate surrounding the question of uniqueness versus permissivism. Although we cannot address this further question here, our result cannot be ignored as lacking a target merely because some proponents of the principle seek to ignore the problem of uniqueness. This is especially so when its originator intended that the 'objective' in 'objective Bayesianism' should distinguish it from subjective Bayesianism's rejection of uniqueness.

I have shown that applying the maximum entropy principle to Bertrand's paradox faces grave difficulties. The first obstruction is damaging because the Bertrand's chord set shows a significant narrowing of the extent to which the maximum entropy function is knowable. This has been neglected because physicists have applied the maximum entropy principle to event spaces with sufficient empirical constraints to be within that narrowed extent. Of course, when you have suitable further information you can avoid the problem

---

[31] For example, take P_1 to be that induced by considering the angle at the vertex of the triangle, P_2 to be that induced by where on the radius the centre of a chord lies. Details in Clark 2002 and Shackel 2007.

here: but this is why I drew attention to the generality of the problem that a claimed solution to Bertrand's paradox faces. The fact that some empirically constrained cases do not fall to this obstruction is not a resolution of it. The second, which I regard as conclusive, shows that the maximum entropy principle falls to Bertrand's paradox or faces a new chord paradox that is structurally similar to, and familiar from, a simple source of paradox for the principle of indifference. The third, a coda based on the supposition that the previous obstructions can be avoided in a principled way, shows that the maximum entropy principle collapses into the original Bertrand's paradox.

In conclusion, then, I have proved that Bertrand's paradox remains the exemplar it has always been. The problem is not confined to probabilities for chords. The obstructions I have shown are a lurking threat for the application of the maximum entropy principle to *any* continuum sized event space lacking a natural measure.

Appendix

I include this appendix (1) for the benefit of those who wanted to check that I haven't concealed philosophically significant moves in the presupposed mathematical results; (2) to demonstrate the status of those presupposed results, including where there are mathematical burdens of proof that a proponent of the MEP[32] faces, some of which I fulfil, or sketch a fulfilment. Despite some difficulties I mention about such burdens, the appendix is not intended as a further line of argument against the MEP. [Could be omitted if editor or referee think the appendix safely dispensable.]

1. We need a definition of continuous entropy

Shannon entropy is defined only for discrete spaces, spaces that are finite or countable. But *C* is continuum sized and since Shannon entropy is not defined for it we need a definition of continuous entropy. Unfortunately there is not a simple definition of continuous entropy that can be considered *the* limiting case of discrete entropy.[33][34] Consequently, scientists have found a variety of definitions of continuous entropy worth using depending on the empirical situation. There is no reason to expect those definitions to agree on the function with maximum entropy, nor to expect the various functions that have maximum entropy under the various definitions to agree on the probability of the longer chord, and hence Bertrand's Paradox recurs immediately.

Nevertheless, there is a definition of continuous entropy, a definition that Shannon offered in an appendix of his original paper (1948:35), which has some claim to be considered standard, namely, differential entropy. It is at least possible that good grounds could be found for this being the uniquely correct definition. For example, if we follow Williamson in seeking the function with minimum Kullback-Leibler divergence from the equivocator function, in finite state spaces this is provably the maximum entropy function (Williamson 2010:28-9). Possibly something similar could be proved for continuous state spaces.[35]

---

[32] Maximum entropy principle.
[33] see Jaynes 1968: 235.
[34] Nor will Kullback's principle of minimum discrimination information, using Kullback-Leibler divergence, help here. Although it is sometimes advanced as a natural extension of the maximum entropy principle to continuous spaces, it only tells us to use the nearest posterior distribution given new information to a *prior* probability distribution, whereas our problem is that we have no prior to start with.
[35] Williamson addresses Kullback-Leibler divergence for continuous spaces at Williamson 2010:154 but does not consider to what extent the result for the finite spaces may carry over.

2. The entropy measure depends on a prior definition of the probability density function

Differential entropy for *C* is given by

$$entropy \ (C_R) = -\int_{c \in C} p(c) \log p(c) d\mu$$

where C_R is a random variable taking values in *C*, *c* is a member of *C*, μ is the probability measure induced on *C* by an application of the Theorem of Induced Σ and μ, and *p* is not a probability function over *C* but a probability density function.[36] The entropy measure for *C* therefore depends on the prior definition of the probability density function *p*. Having induced a measure of *C* and constructed a probability measure, *P*, that measure must now be used to define the probability density function by

$$\int_{c \in \sigma} p(c) d\mu = P(\sigma)$$

where $\sigma \in \Sigma$. So these two integrals together are what are required in order to define the entropy of a probability function induced on *C* by an application of the Theorem of Induced Σ and μ.

3. The integrals need to be well-defined

When we recall that *C* is not a subset of $\mathbb{R}^2$ but of $\mathbb{P}(\mathbb{R}^2)$, we need to check whether the integrals defining the entropy are well-defined. Of course, to prove that they are well-defined is not our burden, but that of those defending the MEP. Nevertheless, I shall now exhibit a plausible route to showing that they are well-defined in standard measure theory.

Recall how Lebesgue integration is defined[37] in terms of a logically prior measure over intervals of $\mathbb{R}^n$: for example, the measure of the interval [*a*,*b*] in $\mathbb{R}$ is |*b-a*|, and of [*a*,*b*]×[*c*,*d*] in $\mathbb{R}^2$ is |(*b-a*)(*d-c*)| (where *a*, *b*, *c* and *d* $\in \mathbb{R}$). Measurable sets are then defined in terms of that measure, then measurable functions in terms of all preimages of all intervals of the function codomain being measurable sets[38], and finally the Lebesgue integral is defined as

$$\int_{\sigma} f d\mu = \sup \left\{ \int_{\sigma} \varphi d\mu : 0 \leq \varphi \leq f, \varphi \ simple \right\}$$

where $\varphi : S \rightarrow \mathbb{R}$ being a simple function means that its range is a finite set of non-negative reals $\{x_1, x_2, \ldots, x_n\}$ for which the preimage, A_i, of each x_i is a measurable set and where for a simple function

$$\int_{\sigma} \varphi d\mu = \sum_{1}^{n} x_i \mu(A_i \cap \sigma)$$

This carries over quite directly to more abstract measure spaces where we are integrating a function, *f*, which takes real values over a member of the relevant sigma algebra to define the integral. That is the case for the integrals defining entropy for *C*. We thereby get the probability density function defined by:

---

[36] Strictly speaking *p* takes members of Σ for its arguments, which are subsets of *C*, so here we define $p(c)=p(\{c\})$.

[37] See, e.g., Capinski and Kopp 2004, Chapter 4.

[38] Given a measurable set, *E*, the function $f:E \rightarrow \mathbb{R}$ is measurable iff for any $I \subseteq \mathbb{R}, f^{-1}(I)=\{x \in: f(x) \in I\}$ is a measurable subset of *E*.

$$P(\sigma) = \int_{c\in\sigma} p(c)d\mu = \sup\left\{\int_\sigma \varphi d\mu : 0 \le \varphi \le p(c), \varphi \; simple\right\}$$

and the entropy

$$entropy(C_R) = -\int_{c\in C} p(c)\log p(c)d\mu$$

$$= \sup\left\{\int_C \varphi d\mu : 0 \le \varphi \le -p(c)\log p(c), \varphi \; simple\right\}$$

On this basis it is reasonable to suppose that the integrals defining differential entropy for C are well-defined, and hence that entropy for C is defined. We shall therefore assume that entropy for C is defined. That being said, if what we have here is correct it establishes nothing more than the existence of the entropies. It is evident that evaluating any such entropies and determining maxima are further non-trivial tasks. Having done this much in aid of the proponents of the MEP I do not feel I am obliged to evaluate any entropies or exhibit any maximal functions.

4. The danger of non-unique probability density functions entailing non-unique entropies

To get the entropy we first require a probability density function to be integrated over, which in turn is implicitly defined by an integral relating it to the probability function, which last is logically prior and has been induced on C by the Theorem of Induced Σ and μ. Absent a uniqueness proof, there is a danger that the probability density function is not *uniquely* defined for each probability function. This would arise if two measures on C induced the same probability function whilst producing distinct probability density functions.

An illustration of the kind of thing I mean is this. Take the state space [0,2] and the sigma algebra $\{\varnothing, [0,1), [1,2], [0,2]\}$. Let $p(x) = \frac{1}{2}$ for x in [0,2] and $p'(x)= 1$ for x in [1/2,3/2] (both zero elsewhere). Then both p and p' agree on the probabilities for events in the sigma algebra but they are distinct probability density functions. Moreover, they give distinct entropies, for p it is log 2 and for p' it is 0. Although in this illustration the state space is discrete and finite, the relevant parallel is when our interest is in a discrete set of events that are determined by an underlying continuous space. Because {longer chords} and [1,2], and {not-longer chords] and [0,1), are equinumerous, there are bijections that we can compound to give a bijection from C to [0,2] to use in the Theorem of Induced Σ and μ.[39] This will then give a probability function for C with two different entropies (although perhaps with a somewhat restricted, if still infinite, algebra of events for C, see potential significance of this below). The example is very simple but it makes clear the general danger. *If* non-uniqueness of obtains for *any* functions on *any* measure space that is the same size as C, then because of the Theorem of Induced Σ and μ, there will be non-uniqueness of this kind for Bertrand's paradox.

Just to be clear of the dangers here, non-uniqueness of probability density function can entail accompanying non-unique entropy for the corresponding probability function. Recalling that the entropy is supposed to measure the presumption of a probability

---

[39] If g is a bijection from {not-longer chords} to [0,1) and h is a bijection from {longer chords} to [1,2], define f:$C{\rightarrow}$[0,2] by $f|_{[0,1)}= g$ and $f|_{[1,2]}= h$. A similar construction for $p'(x)= 1$ for x in [1/2,3/2] and $p''(x)= 1$ for x in [1/4,5/4] foreshadows the return of Bertrand's paradox itself. These pdfs give rise by the same indirect route just articulated to two distinct probability functions on C, P' and P'', having the same entropy and with P'(longer) = ¼ or P''(longer)=½. This specific problem is forestalled by these entropies not being maximal, which is why our the arguments in the body of the paper do not rest on this simple example.

function, non-uniqueness in that measure is itself intolerable for the philosophical reasons motivating the MEP in the first place. This amounts to a subtle, interesting and logically prior occurrence of the kind of non-uniqueness that Bertrand's paradox poses. Bertrand's paradox usually undermines by non-unique probabilities: here it undermines earlier by non-unique entropies themselves being objectionable and by them blocking the route to the MEP furnishing probabilities at all.

If there is a way out of this it is that the algebra of events induced on C (by the Theorem of Induced Σ and μ) in these problematic cases is objectionably restricted. What needs to be done, therefore, is state what are philosophically objectionable restrictions, formulate that condition mathematically and then, under that condition, give a uniqueness proof for the probability density function determined by the formula in section 2 of this appendix.

The burden of providing a uniqueness proof lies on the proponent of the MEP. Nevertheless, on their behalf I offer the following: '

Suppose we have two probability density functions, p and p', defined as above and suppose that for all $\sigma \in \Sigma$, $\int_\sigma p \, d\mu = \int_\sigma p' \, d\mu$. Then by theorem 4.22 (Capinski and Kopp 2004:90), $p = p'$ almost everywhere. It follows that $p \log p = p' \log p'$ almost everywhere. Supposing each of $\log p$ and $\log p'$ is measurable, then each of $p \log p$ and $p' \log p'$ is measurable (Capinski and Kopp 2004:60 theorem 3.5). Consequently by proposition 4.9 (Capinski and Kopp 2004:81)

$$\int_C p \log p \, d\mu = \int_C p' \log p' \, d\mu.$$

and hence

$$entropy \ (C_R) = - \int_C p \log p \, d\mu = - \int_C p' \log p' \, d\mu.$$

So whenever the presuppositions of this proof are satisfied, although we may not have uniqueness of p.d.f.s, their equality almost everywhere ensures uniqueness of entropy.

References

Beck, C. & Schlögl, F. 1993. *Thermodynamics of Chaotic Systems : An Introduction*. Cambridge: Cambridge University Press

Bernoulli, J. 1713. Ars Conjectandi Opus Posthumum. Basileae: Impensis Thurnisiorum,

Bertrand, J. L. F. 1888. *Calcul Des Probabilités*. Paris: Gauthier-Villars et Fils

Capinski, M. & Kopp, P. E. 2004. *Measure, Integral, and Probability*. 2nd ed. London: Springer

Clark, M. 2002. *Paradoxes from a to Z*. London: Routledge

Fishburn, P. C. 1986. The Axioms of Subjective Probability. *Statistical Science*, 1, 335-45.

Greco, D. & Hedden, B. 2016. Uniqueness and Metaepistemology. *Journal of Philosophy*, 113, 365-95.

Gyenis, Z. & Rédei, M. 2015. Defusing Bertrand's Paradox. *The British Journal for the Philosophy of Science*, 66, 349-73.

Jaynes, E. T. 1957a. Information Theory and Statistical Mechanics. *Physical Review*, 106, 620-30.

Jaynes, E. T. 1957b. Information Theory and Statistical Mechanics Ii. *Physical Review*, 108, 171-90.

Jaynes, E. T. 1963. Information Theory and Statistical Mechanics. In *Statistical Physics*. Vol. 3. Ed. Ford, K. W. New York: Benjamin Inc. 182-218.

Jaynes, E. T. 1968. Prior Probabilities. *Systems Science and Cybernetics, IEEE Transactions on*, 4, 227-41.

Jaynes, E. T. 1973. The Well Posed Problem. *Foundations of Physics*, 4, 477-92.

Jaynes, E. T. 1985. Some Random Observations. *Synthese*, 63, 115-38.

Jaynes, E. T. 2003. *Probability Theory : The Logic of Science*. Cambridge: Cambridge University Press

Jost, J. 2015. *Mathematical Concepts*. London: Springer

Keynes, J. M. 1921/2014. *A Treatise on Probability*. London: Macmillan

Khinchin, A. I. 1957. *Mathematical Foundations of Information Theory*. New York: Dover Publications

Laplace, P. S. 1814/1995. *A Philosophical Essay on Probabilities*. Translated by Emory, F. L. and Truscott, F. W. New York: Dover

Paris, J. 2014. What You See Is What You Get. *Entropy*, 16, 6186-94.

Paris, J. & Vencovská, A. 1997. In Defense of the Maximum Entropy Inference Process. *International Journal of Approximate Reasoning*, 17, 77-103.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Revised 2nd printing. ed. San Francisco: Morgan Kaufmann

Rosenkrantz, R. D. 1977. *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*. Dordrecht: Reidel

Salmon, W. C. 1967. *The Foundations of Scientific Inference*. Pittsburgh, PA.: University of Pittsburgh Press

Shackel, N. 2007. Bertrand's Paradox and the Principle of Indifference. *Philosophy of Science*, 74, 150-75.

Shackel, N. MS-a. Disentangling Classical Probabilism, the Principle of Indifference and Bertrand's Paradox.

Shackel, N. MS-b. Epistemic Objectivity, Uniqueness and Objective Bayesianism.

Shannon, C. E. 1948. The Mathematical Theory of Communication. *The Bell System Technical Journal*, 379–423, 623–56. Pages numbers in text refer to the reprint available at the following url: http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

Shimony, A. 1985. The Status of the Principle of Maximum Entropy. *Synthese*, 63, 35-53.

van Fraassen, B. C. 1989. *Laws and Symmetry*. Oxford: Clarendon Press

White, R. 2005. Epistemic Permissiveness. *Philosophical Perspectives*, 19, 445-59.

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