Bertrand's Paradox and the Principle of Indifference.

Nicholas Shackel

Abstract:. The general principle of indifference is supposed to suffice for the rational assignation of probabilities to possibilities. Bertrand advances a probability problem, now known as his paradox, to which the principle is supposed to apply; yet, just because the problem is ill-posed in a technical sense, applying it leads to a contradiction. Examining an ambiguity in the notion of an ill-posed problem shows that there are precisely two strategies for resolving the paradox: the Distinction strategy and the Well-posing strategy. The main contenders for resolving the paradox, Marinoff and Jaynes, offer solutions which exemplify these two strategies. I show that Marinoff's attempt at the Distinction strategy fails and offer a general refutation of this strategy. The situation for the Well-posing strategy is more complex. Careful formulation of the paradox within measure theory shows that one of Bertrand's original three options can be ruled out, but also shows that piecemeal attempts at the Well-posing strategy won't succeed. What is required is an appeal to general principle. I show that Jaynes' use of such a principle, the Symmetry Requirement, fails to resolve the paradox, that a notion of meta-indifference also fails, and that, whilst the Well-posing strategy may not be conclusively refutable, there is no reason to think that it can succeed. So the current situation is this. The failure of Marinoff's and Jaynes' solutions means that the paradox remains unresolved, and of the only two strategies for resolution, one is refuted and we have no reason to think the other will succeed. Consequently Bertrand's Paradox continues to stand in refutation of the general principle of indifference.

Dr N Shackel
James Martin Research Fellow
Faculty of Philosophy and Future of Humanity Institute
University of Oxford
Oxford

nicholas.shackel@philosophy.ox.ac.uk

Bertrand's Paradox and the Principle of Indifference.

Probability and the Principle of Indifference

We can't get numerical probabilities out of nothing. We certainly can't get them out of the mathematical theory of probability, which strictly speaking tells us only what follows from the assumption that certain possibilities have certain probabilities—that is to say, from the assumption of a certain probability measure on a set of possibilities. So before we can apply the theory to a probability problem we have to supply some basis for assuming a particular probability measure.

Classically, this was achieved by determining a base set of mutually exclusive and jointly exhaustive 'atomic' events among which there was no reason to discriminate. These were then assigned equiprobabilities summing to 1. For example, since there are six sides to a die, only one of which can be on top at any one time, the set of atomic events are the six distinct possibilities for which face is on top, each of which is assigned the probability of 1/6. Extending the classical method to the case of infinite sets of possibilities is a bit more complicated. For countable infinities there isn't a way of assigning equiprobability to members of the base set that will sum to 1, but for uncountable infinities there is. Since we will spend quite a lot of time below looking at precisely how it works for uncountable infinities, I won't spell it out now.

The principle being applied here was formulated by J Bernoulli as the Principle of Insufficient Reason, and later by Keynes as

> The Principle of Indifference:…if there is no *known* reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability. (Keynes 1921/1963: 42)

The principle is supposed to encapsulate an a priori truth about the relation of possibilities and probabilities: that possibilities of which we have equal ignorance have equal probabilities. Prima facie, the principle is quite unrestricted. It is supposed to apply to any events or sets of events among which we have no reason to discriminate and to allow equality of ignorance to be sufficient to determine the probabilities.

Bertrand's paradox.

Joseph Louis François Bertrand (1822-1900) was a French mathematician who wrote an influential book on probability theory. In *Calcul Des Probabilités* he argued (among other things) that the principle of indifference is *not* applicable to cases with *infinitely* many possibilities because

> [To be told] to choose *at random*, between an infinite number of possible cases, is not a sufficient indication [of what to do] (1888:4, my translation throughout)

and for this reason to try to derive probabilities in such cases gives rise to contradiction. As proof, he offers many examples, including his famous paradox

> We trace *at random* a chord in a circle. What is the probability that it would be smaller than the side of the inscribed equilateral triangle? (Bertrand 1888:4)

Since subsequent discussion has been in terms of the chord being longer, what I shall from hereon call Bertrand's question is 'What is the probability that a random chord of a

circle is longer than the side of the inscribed equilateral triangle?'. For brevity, I shall speak of the answer to Bertrand's question as the probability of a longer chord. Applying the principle of indifference in three different ways seems to give three different answers:

> (1) The chords from a vertex of the triangle to the circumference are longer if they lie within the angle at the vertex. Since that is true of one-third of the chords, the probability is one-third.

> (2) The chords parallel to one side of such a triangle are longer if they intersect the inner half of the radius perpendicular to them, so that their midpoint falls within the triangle. So the probability is one-half.

> (3) A chord is also longer if its midpoint falls within a circle inscribed within the triangle. The inner circle will have a radius one-half and therefore an area one-quarter that of the outer one. So the probability is one-quarter.
> (Clark 2002:18)

Bertrand concludes that 'the question is ill-posed' (1888:4), and takes it thereby to undermine the principle of indifference — because application of the principle of indifference is supposed to suffice for the entailment of consistent solutions to probability questions, but here it entails contradictory probabilities.

Kinds of ill-posed problem and kinds of solution to Bertrand's paradox.

I know what a problem is, and I think you do too. Consequently we know that problems have identity. However, it is not easy to specify criteria of identity for problems. For example, a problem is not identified by a specification of what counts as a solution because many distinct problems can share the same specification.[1] Nor is it identified by its answer, since many distinct problems can share the same answer. Nevertheless, I have no doubt but that we successful pose and solve problems all the time, and so we have a practical grip on them even if we face difficulties in making that grip theoretically explicit.

By a determinate problem I mean a problem whose identity has been fixed by the way it has been posed. For example, a question which has a single meaning (which singularity might depend not just on the words, but also the context and background constraints) suffices for the problem posed to be determinate. Necessarily, if a problem is determinate then what would count as a solution is determinate. A determinate problem need not have a solution, and if it does, it need not be determinable by us.

In speaking of the determinacy of a problem I might have been speaking of an epistemological matter, a matter of knowing what the problem is or being able to solve it. Certainly, success in fixing the identity of a problem has implications for our epistemic relation to it. Nevertheless, that is not what I am speaking about. The determinacy of a problem is a matter only of its identity — it is a metaphysical matter consequent on facts about the semantics and pragmatics of our ways of expression.

In general, what mathematicians mean by an ill-posed problem is one which requires but lacks a unique solution.[2] There is, however, an ambiguity in the notion of an ill-posed

---

[1] See example shortly.

[2] Mathematicians often speak of ill-posed problems as being ill-posed because they have no or many solutions —a way of speaking which can be objected to since having many solutions is, strictly speaking, a way of not having a solution, that is to say, a way of having no solution as well. However, it can be a matter of mathematical significance whether failure of uniqueness is due to surfeit or surplus, and hence their vocabulary reflects their attention to these distinct modes of failure.

problem. In what is, for our purposes, the primary sense, the fault of ill-posing is the absence of a unique solution to a *determinate* problem. A classical example would be the problem of solving a simultaneous equation when the equations are not linearly independent. Such a problem is determinate and the solution required is a unique tuple of numbers satisfying each of the equations.[3] But linear dependence implies that there are either no or infinitely many tuples that satisfy the equations, and so this problem is ill-posed in the primary sense. This kind of ill-posing is not repairable. Consequently such a problem stands as a refutation of any principle which is supposed to be sufficient (in the context, given the relevant background constraints) for a unique solution.

In the secondary sense, the fault of ill-posing is posing an *indeterminate* problem (whilst nevertheless requiring a unique solution). A problem might be indeterminate because what is to count as a solution has not been determined, but that kind of indeterminacy is irrelevant to Bertrand's Paradox, since what counts as a solution is a unique number in [0,1] being assigned as the probability of the chord being longer. A problem might be indeterminate because as posed it is vague or ambiguous or underspecified. If such an indeterminate problem can be resolved into distinct determinate problems which are well-posed, then this kind of ill-posed problem is no refutation of a principle supposedly sufficient for a unique solution.[4]

Bertrand's Paradox can undermine the principle of indifference if and only if it is ill-posed in the primary sense. If it is ill-posed in the primary sense then it is a determinate probability problem which lacks a unique solution. Yet applying the principle of indifference is supposed to be sufficient for us to solve a determinate probability problem, since such problems have unique solutions.[5] Consequently, the paradox undermines the principle. If Bertrand's Paradox is not ill-posed in the primary sense it is either not ill-posed at all, in which case it doesn't undermine the principle, or it is ill-posed in the secondary sense, i.e. an indeterminate problem. If it is indeterminate, a supporter of the principle of indifference is entitled to sharpen any vagueness and distinguish distinct determinate problems that the question confounds through ambiguity or underspecification. Provided that under such sharpenings and disambiguations the principle suffices for a unique solution to each problem the paradox does not undermine the principle.

Consequently there are two, and only two, different ways of resolving Bertrand's paradox. One way, which I shall call the Distinction strategy, is to concede that it is ill-posed, but to show it to be ill-posed *only* in the secondary sense—by showing that it can be resolved into distinct determinate problems which are not themselves ill-posed in the primary sense. The other way, which I shall call the Well-posing strategy, is to show that it is not ill-posed at all—by showing that it poses a determinate problem for which the principle of indifference is sufficient to determine a unique solution.

---

[3] Continuing from fn. 1, so 'a unique satisfying tuple' is the specification of what counts as a solution that is shared by many distinct simultaneous equation problems.

[4] Distinguishing problems within an indeterminate problem may give rise to a regress, since some of the distinguished problems may themselves be indeterminate. In general, but perhaps not for Bertrand's paradox, showing that the problems to be distinguished within an indeterminate problem are well-posed may require showing that the 'tree' of distinguished problems has no non-terminating branches and that each branch terminates with a determinate well-posed problem. Whilst the 'tree' could be infinitely wide, if it were infinitely deep then the length of the infinite branches would have to correspond to a limit ordinal.

[5] Unique because a solution is a *single* function from the events of interest into [0,1]. The solution being a function entails that no event is assigned two or more probabilities (although distinct events may be mapped to the same probability, of course).

We are going to look at the main contenders in each strategy: Marinoff's use of the Distinction strategy and Jaynes' use of the Well-posing strategy. We will see that Marinoff's solution does not succeed, and that the considerations which undermine it are not specific to his solution but apply to the Distinction strategy as such. We will see that Jaynes' solution, whilst initially attractive, amounts to substituting a restriction of the paradox for the paradox, and hence fails. I shall then show that a notion of meta-indifference (introduced in discussing Marinoff and possibly implicit in some remarks of Jaynes) cannot be used to show the paradox to be well-posed. I shall conclude by summarising the state of play. First, however, I need to formulate the paradox more abstractly than is usually done. I will then be able to show that we have a good reason to reject one of Bertrand's three original ways of assessing the probabilities, before moving on to discussing Marinoff, Jaynes and meta-indifference.

Probability theory

For our analysis we need only the most abstract features of the standard measure theoretic formulation of probability. 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. If A is a σ-algebra on a set S then a measure for A is a non-negative function $\mu : A \mapsto \mathbb{R}$ such that $\mu(\varnothing) = 0$ and μ is countably additive. Countable additivity means that if S_ω is a countable sequence of subsets of S which are pairwise disjoint then $\mu(\bigcup S_\omega) = \sum_{n \in \omega} \mu(S_n)$.[6]

A probability space is an ordered triple $\sqrt{X}, \Sigma, P\cdot$, where X is the sample space of events, Σ is a σ-algebra on X and P is a measure on Σ for which $P(X) = 1$. Being such a measure is sufficient for satisfying Kolmogorov's original axioms (e.g. see Capinski and Kopp 1999:46 Remark 2.6). We shall continue to speak in terms of events, but X can just as well be a sample space of possible worlds or propositions, according to taste.

For completeness, and before moving on, I can now explain two responses to Bertrand's Paradox that are available if one gives up certain views of probability. First, it is possible to avoid the paradox whilst retaining the principle of indifference by allowing finite additivity but denying countable additivity for probabilities. Bertrand himself is arguing for finitism and the finitism got from giving up countable additivity can be motivated independently of his paradox. De Finetti held that 'no-one has given a real justification of countable additivity' (1970:119) and Kolmogorov regarded his sixth axiom (which is equivalent to countable additivity) as needed only for 'idealised models of real random processes' (1956:15). It is true that finitism for probability might, in the end, be a position we have to accept. However, finitism is a severe restriction and may amount to an unacceptably impoverished theory of probability. Furthermore, some philosophers, such as Williamson (Williamson 1999), have been willing to argue contra de Finetti that subjectivists must accept countable additivity. So for good reason we have been unwilling to give in without a fight, and so have continued to try to solve the paradox whilst retaining countable additivity.

The second response can be advanced on the basis of empirical frequentist theories to probability. Defining probability in terms of frequency, and distinguishing reference classes in terms of specifics of empirical situations (for example, a circular flower bed and chords defined by entrance and exit points of overflying birds, a circular container of gas and chords defined by successive collisions with the wall by particles) could well

---

[6] $\bigcup S_\omega$ is the union of all sets in S_ω. S_n is the nth member of S_ω.

result in determinate solutions for such empirical situations. Such solutions might be regarded as examples of the Distinction strategy, and certainly there is no paradox if distinct empirical situations result in distinct probabilities of the longer chord. However, the original point and the continuing importance of the paradox is the challenge it poses to the principle of indifference, and hence to theories of probability that have some reliance on that principle. Frequentist theories *reject* that principle and consequently, frequentist solutions to Bertrand's paradox are somewhat beside the point of the paradox. Indeed, frequentists may advance the paradox as part of an argument against other accounts of probability.

Getting the level of abstraction right

Let C be the set of chords with which we are concerned. In order to calculate the probability of the chord being longer we want to measure the two sets of chords (longer and not longer) and taking the odds to be the ratio between the measures.[7] Setting aside the paradox for the moment, there are other questions to be raised about Bertrand's procedure.

Firstly, only in case (3) is a measure on C itself offered. In cases (1) and (2) what is offered are measures on subsets of C, which subsets are taken to be representative. Why is measuring a subset adequate? Case (2) implicitly partitions C and considers a measure on one equivalence class.[8] Case (1) doesn't partition C since each chord belongs to two such subsets.[9] In both cases the set of similar subsets form a group under the symmetries of a circle and Bertrand explicitly mentions the symmetry fact. This procedure has intuitive geometrical appeal and mathematicians can see how to flesh it out in detail. Bertrand's suggestions for measuring C in the first two cases looks like measuring ratios of an abstract cross section of a measure space which has uniform cross section in order to determine ratios in the whole measure space—rather like measuring the ratio of the volume of pink and white candy in cylindrical seaside rock[10] by measuring the pink and white areas on a slice. If we are not happy with that, well, he has said enough for a mathematician to determine the corresponding measure space he must mean. So Marinoff (1994:5, 7) is misleading us when he represents Bertrand's procedure in these cases as a matter of answering an altogether different problem from that of the chance of getting a longer chord. [11]

---

[7] Odds and probabilities are related thus: the odds of A to ¬A are x:y iff $P(A) = x/(x+y)$ and $P(\neg A) = y/(x+y)$.

[8] In case (2) the relation which partitions C is being parallel. A chord is parallel to itself; if a chord is parallel to another then that other is parallel to it; if a chord is parallel to another and that one to a third, then the first is parallel to the third. Hence the relation of being parallel is an equivalence relation and therefore it partitions C into sets of parallel chords. If the relation to case (2) is unclear, consider partitioning the chords into sets perpendicular to each diameter of the circle. (We can't use the radii mentioned in case (2) because then each diameter belongs to two sets and we don't have a partition.) This gives us the same partitioning into sets of parallel chords that the being parallel relation does.

[9] The subsets are determined for each point x on the circumference of the circle: {c: $c \in C$ and $x \in c$}. Since each chord has two ends there are two such subsets that each chord belongs to.

[10] Seaside rock is a British sweet, a bit like a candy cane, but straight with a pink covering and the name of a seaside resort extruded along the length of the candy so wherever you cut it you see the name. See http://en.wikipedia.org/wiki/Rock_%28candy%29 for a picture.

[11] Furthermore, to be strict we would have to disagree with Marinoff's proposals for the corresponding measure spaces since, for example, his torus on (1994:6) has two points for each member of C, whereas a correct full measure space will have only one point for each member of C.

Secondly, Bertrand equates measures on *C* with measures on $\mathbb{R}$ in the first two cases and a measure on $\mathbb{R}^2$ in the third. What in effect we are being offered is a function from *C* into $\mathbb{R}$ or $\mathbb{R}^2$,[12] and then the Lebesgue measure on the image is taken as a satisfactory measure of *C*. But what is the justification for equating probability measures on *C* with measures on $\mathbb{R}$ or $\mathbb{R}^2$? So far, it is nothing more than an appeal to geometrical intuition and a function between the measured set and the measuring set. We know this can lead us astray when it comes to measure. For centuries mathematicians got into difficulties attempting to use geometrical intuitions and implicit bijections for measuring areas, for example, by 'adding' up the 'lines' from which they were 'composed'.[13] Furthermore, we know that a bijection between sets is insufficient for equality of measure. All line segments have the same cardinality, and hence between any two line segments there exists a bijection, including between line segments of differing lengths. More dramatically, we have the Banach-Tarski theorem, a consequence of which is that a sphere can be decomposed and then recomposed into two spheres of twice the volume. Both being continuum sized entities entails that there is a bijection from the single sphere to the pair of spheres, yet it has half the volume. So the mere existence of a function from *C* into $\mathbb{R}$ or $\mathbb{R}^2$, which function is not even a bijection but which nevertheless captures a certain geometrical intuition, is an inadequate basis for taking a standard uniform measure on $\mathbb{R}$ or $\mathbb{R}^2$ to be a probability measure of *C* got from applying the principle of indifference to *C*. We need, therefore, to investigate more carefully the grounds on which the principle of indifference is applied to continuum sized sets.

Applying the principle of indifference to continua

In continuum sized sets a probability measure cannot be induced by treating members individually. The principle of indifference can only be applied by assigning equiprobability to subsets about which we have a certain equal ignorance: equal ignorance of which subset of events the outcome will belong to. This is done by making use of a uniform measure on those subsets.

---

[12] In case (1), from *C* onto [0, π], in case (2) from *C* onto [-*r*, *r*] and in case (3) from *C* onto $\{(x, y): (x, y) \in \mathbb{R}^2, x^2 + y^2 \leq r^2\}$.

[13] Consider Cavalieri's method of proving the equality of the area of the triangles got from a rectangle by the diagonal. 'If two plane figures have equal altitudes and if sections made by lines parallel to the bases and at equal distance from them are always in the same ratio, then the plane figures are also in this ratio' (Andersen 1985:316). It works, but consider if one dropped the condition of equal distance from the base (and why shouldn't one, since if areas are really constituted by adding up the lines why should their distance from the base make a difference). Then consider a rectangle with a convex curve running from opposite corners. In the latter case the areas are different, yet by the method of comparing the lines from which the area is constituted, they come out the same. Cavalieri succeeded because he found ways round the obvious problems that plagued previous attempts at getting areas from lines, but in doing so he was really leaving behind the geometrical intuitions that were being appealed to in those attempts. On the other hand, we should not forget the success of Newton's geometrical intuition that the height of a curve is the rate at which the area underneath it is increasing, which thought contains the essence of the fundamental theorem of calculus. Again, however, it required the development of analysis in the 19th century before mathematicians stopped committing errors on the basis of this intuition. Measure theory, developed at the turn of the 19th century, is where these problems were finally laid to rest.

For example, consider the case of a random number between 2 and 4. The uniform probability density function for the interval [2,4][14] is an application of the principle of indifference on that basis. Why? Because it amounts to assigning equiprobability to equally long intervals within [2,4]. So in taking the uniform probability density function to be an application of the principle of indifference we are presupposing that the σ-algebra of subintervals of [2,4] is the relevant σ-algebra and we are presupposing the Lebesgue measure on [2,4]. Then, lacking reason to prefer one equally long interval over any other, we take the possibilities of which we are equally ignorant to be the possibilities of the number belonging to equally long intervals, on which basis the principle of indifference entails that equally long intervals should have equiprobability:

For all *I*, *J* that are subintervals of [2, 4], if $L(I) = L(J)$ then $P(x \in I) = P(x \in J)$

Formulated for the general case:

Principle of Indifference for Continuum Sized Sets: For a continuum sized set X, given a σ-algebra, Σ, on X and a measure, μ, on Σ, and given that we have no reason to discriminate between members of Σ with equal measures, then we assign equiprobability to members of Σ with equal measures:

for all x, y in Σ, if $\mu(x) = \mu(y)$ then $P(x) = P(y)$. (This can easily be achieved by setting $P(x) = \mu(x)/\mu(X)$ for all x in Σ.)

I do not know of a plausible non-equivalent way of applying the principle of indifference to continuum sized sets, and so I think this formulation makes clear a requirement for its application to such sets. To apply the principle of indifference to a continuum sized set in order to get a probability measure requires that we are given a σ-algebra on the set and some *other* measure on that σ-algebra relative to which we can assert indifference.[15] Let us call these the required Σ and μ and assume that when given them we apply them to derive the probability measure P by setting $P(x) = \mu(x)/\mu(X)$. We then have the required probability space √X, Σ, P·, where X, Σ and μ were given and P was derived from μ in the way just explained.

Finally we need one further fact. We can use a measure on one set to induce a measure on another.

Theorem of Induced Σ and μ: Given a set, Y, with a σ-algebra, A, and a measure, m, we can use a suitable 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.[16]

Probabilities for C.

In order to apply the principle of indifference for continuum sized sets to C we need a σ-algebra on C to which the set of longer chords, L, belongs, and a suitable measure on that σ-algebra.

---

[14] This p.d.f. is f(x) = ½ for $x \in$[2, 4], and f(x) = 0 elsewhere. Then the probability that the random number is in the interval [a, b] is the area under the graph of f between a and b and that area is ½(b-a).

[15] By which I mean that nothing we know justifies discriminating between members of the σ-algebra which have the same measure, so we have equal ignorance of the possibilities which have equal measure.

[16] I.e. 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 $\mu(S) = m(A)$.

We should be clear that there is no natural measure on C because there is no natural σ-algebra on C which has a measure. We must not let C's close association with $\mathbb{R}^2$ blind us to the fact that C is not a subset of $\mathbb{R}^2$ but of $\mathbb{P}(\mathbb{R}^2)$. σ-algebras of intervals of $\mathbb{R}^2$ and the measures on $\mathbb{R}^2$ are not σ-algebras and measures of C. Consequently to determine probabilities for C we must use the Theorem of Induced Σ and μ in order to apply the Principle of Indifference for Continuum Sized Sets. So we make use of functions from C into measurable sets. But of course, there are infinitely many such functions, so it is likely that for any $x \in [0,1]$ we could find a measure to give us $P(\text{longer}) = x$.

By indicating functions from C into $\mathbb{R}^2$, Bertrand indicates ways of referring to elements of C in terms of their endpoints, or of their centre points.[17] He then passes on to using a uniform measure on the image of C under those functions as if it were a measure of C given by the principle of indifference. However, we need to distinguish ways of referring to the elements of C from ways of measuring C. When we do so it is clear that mere correlations of C with subsets of $\mathbb{R}^2$ don't of themselves justify measuring C in terms of a measure on $\mathbb{R}^2$. What is required is some reason for thinking that the function doing the correlating has some significance for the problem as stated.

Bertrand's argument could therefore be laid out like this:

1. Asking for the probability of a longer chord is a determinate probability problem requiring a unique solution.
2. C is a continuum sized set and lacks a natural measure on which to base a probability measure.
3. So to apply the principle of indifference to C we must induce a measure on C by use of the Theorem of Induced Σ and μ.
4. There are at least 3 mappings from C into $\mathbb{R}^2$ which can be used for inducing the required Σ and μ, and they have equally plausible geometrical significance.
5. But the induced probability measures got from those mappings gives distinct probabilities for the longer chord.
6. Therefore the problem is ill-posed in the primary sense.
7. Therefore the principle of indifference is insufficient to solve this probability problem.
8. But the principle of indifference is supposed to be sufficient to solve all probability problems.
9. Therefore the problem refutes the principle of indifference.

A question can be raised about this conclusion by querying the interpretation of the eighth premiss. In what sense of 'supposed to' is the principle supposed to suffice. Is it supposed to suffice metaphysically, or epistemically? If it is only the former, then it is possible that Bertrand's Paradox doesn't show a failure of the principle of indifference, but rather brings into view a failure of our epistemic capacities. 'The perceptions of some relations of probability may be outside the powers of some or all of us' (Keynes 1921/1963:18). The difficulties we have in getting knowledge of transfinite objects leaves us ignorant of the intrinsic structure of many such objects, and consequently we cannot be

---

[17] I say indicating rather than defining because, as noted above, for ease of exposition he takes the short cut of giving functions from C into $\mathbb{R}$, confident that he can rely on our knowledge of the relevant symmetries to construct the function from C into $\mathbb{R}^2$.

sure that there is not a 'natural' intrinsic measure on *C* in terms of which to define a unique uniform probability function on *C*. So it could be that unbeknownst to us, the principle of indifference does determine a unique probability of the longer chord.

To address this question adequately would require addressing difficult problems in the philosophy of mathematics. For example the claim that transfinite objects have intrinsic structure seems already to be committed to Platonism. This is not the place to address those problems, so I shall make only a few remarks in passing.

I concede that this interpretation of the nature of the problem posed by Bertrand's Paradox might be correct. In that case, the correct conclusion would be only that *we* cannot apply the principle of indifference to *C*. So anyone who thinks that the principle of indifference is a purely metaphysical principle will be entitled to regard Bertrand's Paradox as a rebuttal[18] not of the principle but of our beliefs about the extent of problems to which we are able to apply it.

However, most philosophers of probability who want to make use of the principle of indifference think that objective probabilities are intimately related to rational degrees of belief, and for that reason are unlikely to think that the principle of indifference is a purely metaphysical principle. I said earlier that the principle is supposed to encapsulate an *a priori* truth about the relation of possibilities and probabilities: that possibilities of which we have equal ignorance have equal probabilities. If that formulation, or something close to it, is correct, then the principle of indifference cannot be purely metaphysical and Bertrand's Paradox rebuts the principle of indifference itself.

Excluding case (3)

I now show that we have a reason to exclude Bertrand's case 3, on the basis of general constraints from measure theory.

A null set in a measure space is a set which can be covered by a sequence of other sets whose total measure is arbitrarily small. For example, the rational numbers are null in the real line.[19] Null sets have measure 0. Not all null sets are countable, for example, Cantor's ternary set is null yet uncountable. However, nullity indicates a kind of sparseness within the measure space as a whole, and in general uncountable sets which are not peculiarly constructed, and which are consequently continuous (in a relevant sense) cannot have measure 0.

It seems reasonable to expect applications of the Theorem of Induced Σ and μ to *C* in order to derive probabilities should use bijections for the function, since that would amount to 'counting' *each* chord once and only once.[20] But this restriction is more onerous than it need be. It would not matter if chords that are sparse in *C*, in the sense of sparseness captured by measure theoretic nullity, were not counted at all, or if they got mapped to images which had measure 0. What would be objectionable is if a set of chords in *C* which was not sparse got an induced measure of 0.

The set of diameters, *D*, of a circle is a subset of *C*, and is a continuum sized set. Admittedly there will be *some* bijection which will induce a measure under which *D* is a null set in *C*, just because of the two dimensional Cantor ternary set. However, such a function should not be used for measuring geometrical probabilities on *C* because there is a clear sense in which *D* is not sparse in *C*, just because considered geometrically *D* has a contiguity that amounts to a kind of continuity. D, laid out in terms of that contiguity,

---

[18] I say 'rebuttal' rather than 'refutation' since we are discussing the soundness of my reconstruction of Bertrand's argument.
[19] See Weir 1973:18.
[20] Marinoff 1994:6 counts each chord more than once.

reconstructs the whole disk. *D* is therefore not sparse in *C* and therefore should not get a measure of zero.

Now case (3), when fully spelt out, say for a circle of radius *r* centred on the origin, maps members of *C* onto their midpoints in the disk $\{(x, y): x^2 + y^2 \leq r^2\} \subset \mathbb{R}^2$, and then assigns probability measures for sets of chords on the basis of their area occupied by their midpoints. In so doing, it maps the entire set of diameters onto the origin, a single point and hence null in the disk. Consequently, case 3 amounts to assigning measure 0 to the set of diameters. That is objectionable for the reasons just given, and so case 3 should be ruled out.

Marinoff's rebuttal and the strategy of distinction

Marinoff's 1994 paper has been fairly widely accepted as a successful resolution of Bertrand's paradox. He says

> Bertrand's original problem is vaguely posed … clearly stated variations lead to different but … self-consistent solutions. …[Thus] The principle of indifference appears consistently applicable to infinite sets provided that problems can be formulated unambiguously. (Marinoff 1994:1)

We recognise this as an example of the Distinction strategy. The claim is that the paradox poses a problem whose identity is indeterminate, through vagueness, ambiguity or underspecification, and which can be resolved into a number of distinct determinate problems which are not themselves ill-posed in the primary sense.

Bertrand's question is not vague, since neither having a probability nor being longer than the side of an inscribed triangle are vague properties. Looking at the detail of what Marinoff says, the indeterminacy that he sees in the question is a matter of ambiguity or underspecification. Marinoff accuses Bertrand of failing to specify a random process for selecting the chords:

> When generating random chords, one clearly faces methodological alternatives.…Thus Bertrand's three answers can be construed initially…as replies to three different questions: What is the probability [of a chord being longer] given that the random chord is generated [by a procedure]
>
> Q1 …on the circumference of the circle?
>
> Q2 …outside the circle?
>
> Q3 …inside the circle? (1994:4)

By the end of the paper, Marinoff has distinguished an additional four such questions, giving seven in all, and allows that there may be 'an infinite number'(1994:17). So Marinoff's position is that Bertrand's question confounds distinct problems. What is Marinoff's argument? He doesn't give one, but quotes Keynes and van Frassen approvingly:

> Keynes concludes. "So long as we are careful to enunciate the alternatives in a form to which the Principle of Indifference can be applied unambiguously, we shall be prevented from confusing together distinct problems, and shall be able to reach conclusions in geometrical probability which are unambiguously valid" (Marinoff 1994:23)[21].

---

[21] I have used double quotation marks to distinguish what Marinoff is quoting from Keynes (and later, van Frassen) from what he is saying himself. The Keynes quotation is from Keynes 1921/1963:63.

Response: This study has endeavoured to follow Keynes's positivistic prescription. Careful enunciations of alternatives, unambiguous applications of the principle of indifference, and clear demarcation between distinct problems together lead to conclusions in geometric probability that are self-consistent and therefore unparadoxical. (Marinoff 1994:23)

"Most writers commenting on Bertrand have described the problems set by his paradoxical examples as not well posed. In such a case, the problem as initially stated is really not one problem but many. To solve it we must be told *what* is random; which means, *which* events are equiprobable; which means, *which* parameter should be assumed to be uniformly distributed." (van Frassen 1989:305 as quoted in Marinoff 1994:4-5)

Marinoff states that he is 'implementing van Frassen's recommended method' (Marinoff 1994:4-5), which is odd, since immediately following the passage quoted by Marinoff van Frassen makes an objection:

But that response asserts that in the absence of further information we have no way to determine the initial probabilities. In other words, this response rejects the Principle of Indifference altogether. After all, if we were told as part of the problem which parameter should receive a uniform distribution, no such Principle would be needed. It was exactly the function of the Principle to turn an incompletely described physical problem into a definite problem in the probability calculus. (van Frassen 1989:305)

Marinoff offers no response to this objection. One point available to him is that giving a parameter a uniform distribution is itself an application of the principle of indifference, so his approach is not a rejection of that principle altogether. But what will he say about van Frassen's final sentence?

For the sake of argument, grant that Marinoff's Q1, Q2 and Q3 (and his others) are well-posed distinct problems to which the principle of indifference can be applied successfully. If his object was the restricted one of making plausible the application of the principle in some infinite cases then he may have succeeded. Certainly, that is a strong rebuttal of Bertrand's finitistic rejection of probabilities for any infinite cases.

Significantly, the many versions of Bertrand's problem are solvable, and each solution relies upon the very procedure—namely, the consistent application of the principle of indifference to infinite sets—that Bertrand proscribed. Bertrand's former paradox of random chords is resolved by the expedient of providing what he, from the outset, withheld, namely, a "sufficient specification" of such sets. (Marinoff 1994:22)

But does distinguishing these several problems really get to grips with Bertrand's broader challenge? Bertrand might concede his finitism and still hold that his question embarrasses the principle of indifference by confronting us with distinct but contradictory ways of applying the principle to a single problem.

Hence the bone of contention is whether Bertrand's question poses a determinate problem which lacks a unique answer, or poses an indeterminate problem which through ambiguity or underspecification confounds distinct determinate problems. For Marinoff's resolution to succeed he must persuade us that Bertrand's question is of the latter type. Marinoff wants to be able to reply that if by choosing randomly you mean process p, then the probability is x, but if by choosing randomly you mean process q, then the probability

is *y….*, and if you don't specify what you mean by choosing randomly, then you haven't posed a determinate problem. For

> There exists a multiplicity, if not an infinite number, of procedures for generating random chords of a circle. The answers that one finds to Bertrand's generic question … vary according to the way in which the question is interpreted, and depend explicitly upon which geometric entity or entities are assumed to be uniformly distributed. (Marinoff 1994:17)

Now if Bertrand's question is a generic singular question like 'what is the weight of Fred', then the question can be rejected as underspecified if no particular Fred is contextually salient and the asker refuses to identify which Fred 'Fred' stands for. If he goes on to say that his question is a general question, that he is interested in the weight of Freds in general, it can be rejected as meaningless. Weights are properties of material individuals, but there are no such individuals as Freds in general. (I shall consider later what role the notion of the weight of the average Fred might play).

Marinoff's solution requires that Bertrand's question be similarly and *only* a generic singular question. But what are the grounds for insisting that Bertrand is confined to speaking of random choice in the singular when asking about chords chosen at random? Of course, Bertrand *could* ask about the chance of getting a longer chord when a particular way of choosing randomly is salient, but what he is interested in knowing is what is the chance of getting a longer chord given random choice *in general*. Marinoff would like to reject the general question, but the analogy doesn't carry through because, whilst there is no such thing as a Fred in general there is such a thing as random choice in general.

Furthermore, that a question has several answers doesn't of itself mean that distinct problems are being confounded. When neither the financial institution nor the riverside is contextually salient, the question 'how can I get to the bank' leaves it indeterminate which problem is being posed. But if we know that the riverside is the goal, then there being different ways to get to the riverside doesn't mean that the question is confounding several distinct problems. It is just a single problem with several solutions.

Bertrand's question is not analogous to the former example, but to the latter. We know perfectly well what question has been asked. He wants to know the chance of getting a longer chord. What is it about there being different ways choosing at random which justifies taking his question to be confounding distinct problems which doesn't make the question of the way to the bank similarly confused? Without a good reply to that challenge, I do not see how Marinoff has resolved the paradox.

On the contrary, Bertrand's point seems very well taken. The principle of indifference is supposed to deal with what is unknown by validating the application of indifference over the equally unknown. By choosing a set which lacks a natural measure relative to which equiprobability can be assigned he exposes the significance of that relativity for the general application of the principle in infinite cases. If there is a well motivated restriction on that relativity, such as may be given by a question which gives more information, all well and good. But if there isn't such a restriction, there doesn't seem to be a principled way to get out of the difficulty Bertrand's Paradox poses. Not knowing which way of choosing a chord at random is to be used shouldn't be a problem, since ignorance is what the principle of indifference is supposed to allow us to deal with. Indeed, ignorance is its foundation. But if ignorance does not give reason to discriminate between distinct ways of applying the principle, and if those ways result in contradictory probabilities for the same event, the principle has failed.

The points I have just made do not apply only to Marinoff, but apply to any Distinction strategy. Any such strategy requires as a basic premiss that Bertrand's question is a generic singular question which cannot be a general question. But there seem to be no grounds for rejecting it as a general question. It seems just as meaningful to ask the question in the light of random choice in general as in the light of a particular method of random choice. Furthermore, as van Frassen pointed out, if we are told which method of random choice to use we don't need the principle of indifference, so this strategy is in danger of merely evading the challenge Bertrand's Paradox poses to that principle.

Secondly, even if the rejection of the general question could be maintained, generalising statistically over a generic singular question is itself a procedure warranted by the principle of indifference. For example, although there is no such thing as a Fred in general, the principle of indifference warrants the statistical notion of the weight of the average Fred. If we are ignorant of which method of random choice has been used, that is just more ignorance, and so equiprobability should be assigned to *those* possibilities. I'm going to call this *meta-indifference*. I shall discuss meta-indifference at greater length later, and here I shall consider only what bearing the (epistemic) possibility of meta-indifference has on the Distinction strategy.

Either consistent numerical probabilities can be derived by the application of meta-indifference to Bertrand's Paradox or they cannot. If they can we find ourselves with a unique answer to the statistically generalised generic question in the same sense that the weight of the average Fred is a unique answer to the statistically generalised question of the weight of Fred. However, in that case, we have not vindicated the Distinction strategy, but the Well-posing strategy, for we have shown Bertrand's Paradox to be well-posed in the sense that the principle of indifference is sufficient to turn a generic underspecified question into a determinate statistically general problem with a unique solution.

If meta-indifference doesn't entail consistent numerical probabilities, either it entails inconsistent numerical probabilities or it fails to entail any probabilities. If the former, Bertrand's Paradox has recurred at the meta-level. If the latter, the supporters of the Distinction strategy may feel vindicated. They may argue as follows: So long as there seemed to be a viable notion of the probability of a longer chord in general, even just the etiolated sense got from the statistical generalisation of the generic question, that possibility could be held up as a reproach to our strategy. However, just as there is no such thing as a Fred in general, the failure of meta-indifference to entail any probabilities *proves* that there is no such thing as a probability of the longer chord in general, not even in the etiolated sense. To demand that the principle of indifference be sufficient to calculate a probability *that does not exist* is no reproach. Consequently, all that there can be are the distinct particular problems into which we analyse Bertrand's generic question. Since the principle of indifference is sufficient to solve those problems, it is untroubled by Bertrand's Paradox.

I am unpursuaded that the failure of meta-indifference to entail the probability of a longer chord in the statistically general sense means there is no such probability. Even if we granted (which I do not) that there is no general Bertrand's question except for the statistical generalisation of the generic question, that statistically general question poses a determinate statistically general problem which the principle of indifference is supposed to suffice to solve. That meta-indifference fails to entail the probability of a longer chord does not, for me, prove the non-existence of the probability but the failure of the sufficiency of the principle of indifference.

I would concede, however, that having pushed the argument this far, the matter is finely balanced, and my opponent has further resources to deploy. For example, he might argue

that being a *statistical* generalisation entails that a criterion of identity for the relevant probability is that meta-indifference suffices to calculate it (and so, since it doesn't suffice, the probability doesn't exist). During my discussion of meta-indifference below I shall be offering my reasons for thinking that it *cannot* fail to entail the probability of a longer chord.

So the basic premiss of the Distinction strategy (that Bertrand's question is a generic singular question which cannot be a general question) is probably false, when the strategy must fail. Even if it is true, the principle of indifference warrants being indifferent over the distinctions to be made between the instances of a generic question (because it warrants the relevant statistical concepts). Only if the premiss is true *and* meta-indifference fails to entail a probability of a longer chord does the strategy have any prospects. But even then, that failure may be as much a reproach to the principle of indifference as succour to the strategy. If, as I shall argue below, meta-indifference *cannot* fail to entail the probability of a longer chord, we should conclude that the Distinction strategy can never succeed in resolving Bertrand's Paradox.

Jaynes' rebuttal and the strategy of well-posing

Jaynes agrees that Bertrand's question is general, and argues that its very generality furnishes invariance constraints on acceptable probability measures of *C*.

> If we start with the presumption that Bertrand's problem has a definite solution *in spite of the many things left unspecified,* then the statement of the problem automatically implies certain invariance properties. (Jaynes 1973:480)

He is going to apply what van Frassen calls

> the great Symmetry Requirement: problems which are essentially the same must have essentially the same solution (van Frassen 1989:259)

Jaynes' point is that Bertrand is asking about circles in general, not about particular circles, and so any acceptable probability measure on the set of chords must not depend on accidents of position and scale of the circle concerned, but should rather be invariant over those accidents.

> If…the problem is to have any definite solution at all, it must be "indifferent" to ….small changes in the size or position of the circle. This seemingly trivial statement… fully determines the solution. (Jaynes 1973:480)

Jaynes motivates the symmetry requirements by referring to a 'tossing straws onto a circle' (Jaynes 1973:478) and how this empirical situation should give the same results for distinct observers, that is to say, for distinct frames of reference, for whom the circle may appear rotated, scaled or translated relative to each other. This is why he speaks in terms of small changes. However, examination of his mathematics does not make it evident that his result is valid only for small changes. It appears that his probability measure is quite generally rotationally, scale and translationally invariant. Furthermore, it turns out that

> the requirement of translational invariance is so stringent that it already determines the result uniquely (Jaynes 1973:485)

The mathematical problem as Jaynes sets it up is this:

> The position of the chord is determined by giving the polar coordinates (r, θ) of its center. We seek to answer a more detailed question than Bertrand's:

> What probability density $f(r, \theta)dA$…should we assign over the interior area of the circle? (Jaynes 1973:481)

Jaynes is going to determine a probability density function *not* directly on the set of chords but over the disk $\{(x, y): x^2+y^2 \leq R^2\} \subset \mathbb{R}^2$. The probability measure will be induced on C by the mapping from C to the disk $\{(x, y): x^2+y^2 \leq R^2\} \subset \mathbb{R}^2$ which maps each chord onto its midpoint.

So Jaynes is using the Theorem of Induced Σ and μ. But he is not using a uniform measure on that disk as a way of applying the principle of indifference. Rather, he is applying it like this. Take a small area, Γ, in one circle and the subset of chords, S, picked out by having their centres in that set. Then consider an offset circle and the set of chords, S', in that circle that are collinear with a chord in the first set. Their centres define a small area in the second circle, Γ'. The collinearity of chords defines a bijection between these two areas, Γ and Γ'. The principle of indifference is applied by

> assign[ing] equal probabilities to the regions Γ and Γ', respectively, since (a) they are probabilities of the same event, and (*b*) the probability that a straw which intersects one circle will also intersect the other, thus setting up this correspondence, is also the same in the two problems. (Jaynes 1973:484)

There is a unique density function which possesses this translational invariance:

$f(r, \theta) = 1/(2\pi R r)$, $0 \leq r \leq R$, $0 \leq \theta \leq 2\pi$ (Jaynes 1973:485)

Since $\int f(r, \theta)dA = \int f(r, \theta)\ r\ dr\ d\theta = \int 1/(2\pi R)\ dr\ d\theta$ we find the probability that the chord is longer (i.e. its centre is in the circle inscribed within the inscribed equilateral triangle) is ½.

We recognise Jaynes' solution as an example of the Well-posing strategy. Jaynes denies that Bertrand's Paradox is ill-posed at all, and asserts that it poses a determinate problem for which the principle of indifference is sufficient (given the relevant background constraints) to determine a unique solution. What should we make of this?

van Frassen seems to accept Jaynes' solution for Bertrand's paradox, but draws attention to Jaynes' concession that he doesn't see how to apply his approach to Von Mise's water and wine problem. (van Frassen 1989:315). I think van Frassen is too sanguine.

Marinoff thinks that Jaynes' has fallen into the trap of disputing

> which of these questions [Marinoff's Q1, 2 or 3]—if any—"best" represents Bertrand's generic question. (1994:21)

Marinoff thinks that Jaynes has given an answer to Marinoff's Q2, partly because they seem to agree on the answer to the empirical case of straw tossing. However, this is a significant misrepresentation of what Jaynes is doing. What Jaynes is doing is far more sophisticated, and strictly speaking, they do not agree on the answer to the empirical case.

Marinoff's Q2 is specified in terms of 'the random chord generated…by a procedure outside the circle'(1994:4). Marinoff's solution to Q2 (1994:7-11) is that the probability of the longer chord = the *limiting* probability as the distance of a point outside the circle from the centre of the circle tends to infinity, and that limit = ½. If we are to understand Marinoff's Q2 as relevant to the empirical case, we must understand him as construing the procedure outside the circle as follows: the centre of the straw determines a point outside the circle. The length of the chord generated depends on the angle the straw makes with the extended diameter intersecting the centre of the straw. The angle is assumed to have a uniform probability density function, and straws which do not intersect the circle are ignored.

16

Marinoff is mistaken when he takes his limiting procedure to give the solution to the empirical case. Rather, it approximates cases where the straws are long relative to the circle diameter, so that the chance of the centre of the straw lying inside the circle is negligible. Furthermore, his solution method should not be taken to the limit for cases of specific relatively long straws. When it is not, for such relatively long straws of specific lengths his solution method will result in the probability of a longer chord being *strictly less* that ½, whereas Jaynes' solution to cases of finite straw length is precisely ½.

Marinoff's solution is a mere approximation in a restricted range of empirical cases because Marinoff's solution method to Q2 excludes straws whose centre lies within the circle, whereas Jaynes is exactly correct for an unrestricted range of empirical cases precisely because his solution does not exclude those straws. That the probabilities in Marinoff's solution converge quickly to ½ as straw length increases disguises this important distinction between their solutions.

Now for Marinoff, idealising to straws of infinite length gets rid of the problem of ignoring the straws whose centres are inside the circle. But that really amounts to abandoning the notion of a 'random chord generated…by a procedure outside the circle'. Instead, it turns out that talk of points on extended diameters, uniform distributions over angles of lines through such points and taking limits as the distance of that point from the circle tends to infinity is an obscure way to obtain the answer for the probability of longer chords got from randomly selected lines (rather than line segments) in the plane. But put baldly like that, one now awaits a justification for why the former process should be taken as a solution to the latter problem.

Jaynes faces none of these problems, and in fact, only his approach can satisfactorily explain why the idealisation of infinite straws might be an answer to chords got from randomly selected lines in the plane. Jaynes' mathematics can apply to line segments (straws of specific lengths) but is independent of the finitude of such line segments. Consideration of the way he is applying the principle of indifference in terms of regions Γ and Γ' makes it clear that nothing *depends* on the relevant circles being close (as they have to be for finite straws), but allows them to be arbitrarily distant. In effect, his solution concerns itself with invariance of probability measure *given* infinite lines. This is significant, since it is the reason I think Jaynes' attempt at demonstrating the problem to be well posed fails; it is why Marinoff's criticism, although based on mistaking the relation of Jaynes' solution to Marinoff's Q2, is correct insofar as he convicts Jaynes of solving a particular version of 'Bertrand's generic question' rather than the general question.

The problem is that if we don't accept the fully general mathematical extension of the empirical situation, then the problem Jaynes is considering is not Bertrand's, but a restriction of Bertrand's, not exactly Marinoff's Q2, but a restriction all the same, namely of a process of random choice relative to finite lines in the plane. If on the other hand we accept the full generalisation, the situation is not improved. For quite clearly, what his application of the principle of indifference relies on is families of infinite lines which coordinate many regions Γ, Γ', Γ'', Γ'''…. in many circles. Now this is indeed well motivated for the empirical problem of straw tossing but not for the general problem, since it still counts as specifying a particular way to select chords, namely, selecting them relative to infinite lines in the plane. Bertrand's question is about any circle, not about any circle such that if this chord is selected in this circle, then that collinear chord is selected in that circle, and so on for all circles intersected by the extension of the chord in the first circle. Nothing about the problem as stated, nothing about the generality of circles spoken of, requires *this* coordination of events. It is rather the empirical situation of straw tossing

that does so. That is to concede that Bertrand's general problem has not been solved, but only a particular problem, namely, the idealisation of the straw tossing variant. Jaynes claims that 'we do no violence to the problem if we suppose we are tossing straws' (Jaynes 1973:478), but it turns out that we do.

Meta-indifference

I discussed above Marinoff's claim that Bertrand's question is a generic question rather than a general question. I argued that even if Marinoff *could* reject the general question, the principle of indifference warrants a statistical generalisation over the generic questions. Ignorance of which method of random choice has been used is just more ignorance, and the principle of indifference is supposed to warrant applying equiprobability over equal ignorance, so equiprobability should be assigned to *those* possibilities. I called this *meta-indifference*. I then showed that the Distinction strategy fails unless Bertrand's question is a generic question which cannot be a general question *and* meta-indifference fails to entail a probability of a longer chord. I shall now argue that meta-indifference cannot fail to entail a probability of a longer chord. That rules out the Distinction strategy.

Meta-indifference might entail consistent probabilities for the probability of the longer chord, and were it to do so meta-indifference would thereby provide a solution to Bertrand's Paradox by the Well-posing strategy. I shall also show that Bertrand's Paradox recurs for meta-indifference, and so meta-indifference also fails to solve the paradox.

Before I give those arguments, I concede that there could be other notions of meta-indifference. We can, perhaps, see one such glimmering in Jaynes' broader conclusion:

> it is dangerous to apply this principle at the level of indifference between *events,* because our intuition is a very unreliable guide in such matters, as Bertrand's paradox illustrates. However, the principle of indifference may, in our view, be applied legitimately at the more abstract level of indifference between *problems*; because that is a matter that is definitely determined by the statement of a problem, independently of our intuition. (1973:488)

It is noteworthy that despite what he says here, in his solution to Bertrand's paradox he makes use of indifference between problems to determine events over which to be indifferent (see the quotation from p. 484 above). Nevertheless, I think Jaynes' notion of indifference over problems could be taken to be a kind of meta-indifference. Perhaps there are many ways to be meta-indifferent and perhaps a variety of notions of meta-indifference could usefully be developed. Among those there might be some which avoid the arguments I shall shortly make. The question would then be whether they do any useful work in addressing Bertrand's paradox.

With that caveat in mind, it seems to me that an essential part of any notion of meta-indifference that *could* do work in addressing Bertrand's paradox is that ignorance of method of random choice implies being indifferent over probability measures. I shall now formulate meta-indifference abstractly on that basis, and develop its implications.

> *The Principle of Meta-indifference*: Given a sample space of events, X, a σ-algebra, Σ, on X, a set of probability measures on Σ, M, and given that we have no reason to discriminate between members of M, then we assign equiprobability to members of M and calculate probabilities thus:
>
> for all x in Σ, $P(x) =$ the mean over all μ in M of $\mu(x)$

Implicit in this definition is the treatment of M itself as a probability space (i.e. there being an ordered triple $\sqrt{M}, \Sigma_M, P_M \cdot$, where Σ_M is a σ-algebra on M and P_M is a measure

on Σ_M for which $P_M(M) = 1$). Unless there was a natural uniform measure on M (strictly speaking, on Σ_M), determining equiprobability on M would have to make use of the definitions given above of the Principle of Indifference for Continuum Sized Sets (extended to relevantly sized sets) and also the Theorem of Induced Σ and μ. Clearly, if there is more than one measure on M, and there is no measure which makes sense as *the* uniform measure, Bertrand's Paradox recurs because of distinct and contradictory ways of assigning equiprobabilities to members of M.

In the case of the set of chords, C, we have the set of probability measures on C, M_c, being treated as a probability space. Because the individual members of M_c are measurable (Edwards 1995:198) I believe that there is no difficulty in constructing probability functions on M_c from those measures. For example, there are standard ways of applying integration to measure 'distances' between measurable functions which might be used in such a construction. Consequently meta-indifference cannot fail to entail a probability of the longer chord, and so the Distinction strategy must fail.

Furthermore, M_c is itself quite as abstract as C, and consequently lacks a natural measure in terms of which to use the Theorem of Induced Σ and μ to define equiprobability for members M_c. But that means that for every measure on M_c (of which there are infinitely many) there will be a distinct way of constructing equiprobability for members of M_c, and some of those ways will be contradictory. Therefore Bertrand's Paradox recurs at the meta-level and I think it is clear that this leads to a vicious regress. So meta-indifference cannot be used to give a solution to Bertrand's Paradox by the Well-posing strategy.

Conclusion

Examining the ambiguity in the notion of an ill-posed problem brought into view the only two strategies for resolving Bertrand's paradox: the Distinction strategy and the Well-posing strategy. The main contenders for resolving the paradox, Marinoff and Jaynes, offer solutions which exemplify these two strategies. Our analysis of Marinoff's attempt at the Distinction strategy led us to a general refutation of this strategy. The situation for the Well-posing strategy is more complex.

Formulating the paradox precisely showed that one of Bertrand's original three options can be ruled out, but also showed that piecemeal attempts at the Well-posing strategy won't succeed. There are *continuum* many options and what is therefore required is an appeal to principles sufficient for restricting that many options.

I have not proved that the Symmetry Requirement cannot underpin a successful attempt at the Well-posing strategy. Nor have I proved that there are no other principles that might underpin a successful attempt at the strategy. Unless someone can see a way of deriving a contradiction from the assumption that this strategy can succeed, I doubt if this strategy can be conclusively refuted.

However: I have proved that one principle, meta-indifference, cannot underpin a successful attempt at the Well-posing strategy; and meta-indifference failing due to recurrence shows that if the strategy can work at all, it can work at the base level. Jaynes' attempt at the strategy is a very sophisticated use of the Symmetry Requirement, and yet it turned out to amount to substituting a restriction of the problem for the general problem. I do not know of any other principles that look as if they might help. There is, so far as we can know, no 'natural' measure on the set of chords. No one has succeeded in justifying the claim that any particular measure on the set of chords is the correct measure for the general problem. All in all, then, there is no reason to think that the Well-posing strategy can succeed

So the situation is this. The failure of Marinoff's and Jaynes' solutions means that the paradox remains unresolved. The Distinction strategy is refuted and there is no reason to think that Well-posing strategy can succeed. Consequently, if we wish to retain countable additivity in probability, Bertrand's Paradox continues to stand in refutation of the general principle of indifference.

References

Andersen, K. 1985. Cavalieri's Method of Indivisibles. *Archive for History of Exact Sciences*, 31 pp. 291-367.

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

Capinski, M. & Kopp, P. E. 1999. *Measure, Integral, and Probability*. London: Springer.

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

de Finetti, B. 1970. *Theory of Probability Vol. 1*. New York: Wiley.

Edwards, R. E. 1995. *Functional Analysis*. New York: Dover Publications.

Jaynes, E. T. 1973. The Well Posed Problem. *Foundations of Physics*, 4 (3), pp. 477-92.

Keynes, J. M. 1921/1963. *A Treatise on Probability*. London: Macmillan.

Kolmogorov, A. N. 1956. *Foundations of the Theory of Probability*. 2nd English ed. New York: Chelsea.

Marinoff, L. 1994. A Resolution of Bertrand's Paradox. *Philosophy of Science*, 61 (1), pp. 1-24.

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

Weir, A. 1973. *Lebesgue Integration and Measure*. Cambridge: Cambridge University Press.

Williamson, J. 1999. Countable Additivity and Subjective Probability. *British Journal for the Philosophy of Science*, 50 (3), pp. 401-16.